

# CONVEX BILLIARDS ON CONVEX SPHERES

PENGFEI ZHANG

ABSTRACT. In this paper we study the dynamical billiards on a convex 2D sphere. We investigate some generic properties of the convex billiards on a general convex sphere. We prove that  $C^\infty$  generically, every periodic point is either hyperbolic or elliptic with irrational rotation number. Moreover, every hyperbolic periodic point admits some transverse homoclinic intersections. A new ingredient in our approach is Herman's result on Diophantine invariant curves that we use to prove the nonlinear stability of elliptic periodic points for a dense subset of convex billiards.

## 1. INTRODUCTION

The dynamical billiards, as a class of dynamical systems, were introduced by Birkhoff [Bir17, Bir27] in his study of Lagrangian systems with two degrees of freedom. A Lagrangian system with two degrees of freedom is isomorphic with the motion of a mass particle moving on a surface rotating uniformly about a fixed axis and carrying a fixed conservative field of force with it. If the surface is not rotating and the force vanishes, then the particle moves along geodesics on the surface. If the surface has boundary, then the resulting system is a billiard system. Dynamical billiards on curved surfaces is related to the study of quantum magnetic confinement of non-planar 2D electron gases (2DEG) in semiconductors [FLBP], where the effect of varying the curvature of the surface corresponds to a change in the potential energy of the system. The dynamical billiards can be viewed as a mathematical model for this system, and may be used to investigate the electron transport properties of the semiconductors. As mentioned in [GSG99], the advances in semiconductor fabrication techniques allow to manufacture solid state (mesoscopic) devices where electrons are confined to curved surfaces.

The classical results of dynamical billiards are closely related to geometrical optics, which has a much longer history. For example, the discovery of the integrability of elliptic billiards, according to Sarnak [Sar11], goes back at least to Boscovich in 1757. Surprisingly, the billiard dynamics is also related to the spectra property of Laplace–Beltrami operator on manifolds with a boundary. More precisely, Weyl's law in spectral theory gives the first order asymptotic distribution of eigenvalues of the Laplace–Beltrami operator on a bounded domain. Weyl's conjecture on the second order asymptotic distribution was proved by Ivrii [Ivr80] for any compact manifold with boundary, under the assumption that the measure of periodic points of billiard dynamics on that manifold is zero.

Current study of dynamical billiard systems mainly focuses on the Euclidean case. Birkhoff studied the dynamical billiards inside a convex domain on the plane. Birkhoff also conjectured that ellipses are the only integrable billiards. A weak version of this conjecture was proved by Bialy [Bia93]. The dynamical billiards on a bounded domain with convex scatterers were introduced by Sinai in his study of Boltzmann Ergodic Hypothesis [Sin70] on ideal gases. Sinai discovered the dispersing mechanism and proved that dispersing billiards are hyperbolic and ergodic. Since then, the mathematical study of chaotic billiards has developed at a remarkable speed, and the defocusing mechanism for chaos were discovered by Bunimovich [Bun74, Bun92], Wojtkowski [Woj86], Markarian [Mar88] and Donnay [Don91]. Very recently, the dynamics of some asymmetric lemon billiards are proved to be hyperbolic [BZZ14], for which the separation condition in the defocusing mechanism was strongly violated. See [Vet84, KSS89, GSG99] for the study of chaotic billiards on general surfaces. The study of chaotic billiards also provides the key idea for the construction of hyperbolic geodesic flows on  $S^2$ , see [Don89, BuGe89].

---

2000 *Mathematics Subject Classification.* 37D40, 37D50, 37C20, 37E40.

*Key words and phrases.* convex sphere, convex billiards, generic properties, Kupka–Smale, hyperbolic periodic point, Poincaré's connecting problem, homoclinic intersection, elliptic periodic point, Moser stable, Diophantine, prime-end.

In this paper we consider the convex billiards on convex spheres. Recall that a Riemannian metric  $g$  on the 2D sphere  $S^2$  is said to be (strictly) *convex*, if it has positive Gaussian curvature:  $K_g(x) > 0$  for all  $x \in S^2$ . Given a tangent vector  $\mathbf{v} \in T_x S^2$ , the geodesic passing through  $x$  in the direction of  $\mathbf{v}$  is defined by the exponential map  $\gamma_{\mathbf{v}} : \mathbb{R} \rightarrow S^2$ ,  $t \mapsto \exp_x(t\mathbf{v})$ . For any two points  $p, q \in S^2$ , let  $d(p, q)$  be the length of the shortest geodesics connecting  $p$  and  $q$ . Let  $\text{Inj}(S^2, g)$  be the injective radius of  $(S^2, g)$ .

**Example.** Let  $S^2$  be the unit sphere in  $\mathbb{R}^3$  endowed with the round metric  $g_0$ . Then  $K_0 \equiv 1$ , and every geodesic on  $S^2$  moves along a great circle. Let  $p, q \in S^2$  be two points on the sphere, and  $\alpha$  be the angle between the two position vectors  $\mathbf{p}, \mathbf{q}$ . Then the geodesic distance  $d_0(p, q)$  between  $p$  and  $q$  is given by  $d_0(p, q) = \alpha(\mathbf{p}, \mathbf{q})$ , and  $\cos \alpha = \langle \mathbf{p}, \mathbf{q} \rangle$ . Therefore,  $d_0(p, q) = \arccos \langle \mathbf{p}, \mathbf{q} \rangle$ . Moreover,  $\text{Inj}(S^2, g_0) = \pi$ . The dynamical billiards inside a convex domain of spheres with constant curvatures have been studied recently in [Bol92, Bia13, CP14].

**Definition 1.** Let  $(S^2, g)$  be a convex sphere. A closed subset  $Q \subset S^2$  is said to be (geodesically) *convex*, if  $Q$  is simply connected, and for any two points  $x, y \in Q$ , there is a unique minimizing geodesic contained in  $Q$  connecting  $x$  and  $y$ . A convex domain  $Q$  is said to be *strictly convex*, if the interior of each minimizing geodesic is contained in the interior  $Q^\circ$  of  $Q$ .

Let  $Q \subset S^2$  be a convex domain,  $s$  be the arc-length parameter of  $\Gamma = \partial Q$ , and  $\kappa(s)$  be the geodesic curvature of  $\Gamma$  at  $\Gamma(s)$ . Note that  $\kappa(s) \geq 0$  for all  $s$ . If  $Q$  is strictly convex, then  $\kappa(s) > 0$  for all  $s$  (except on a closed set without interior). By definition, there are no conjugate points inside a convex domain  $Q$ . In the following we require that there are no conjugate points on the closed domain  $Q$ . A sufficient condition for nonexistence of conjugate point is that  $\text{diam}(Q) < \text{Inj}(S^2, g)$ .

The dynamical billiard on  $Q$  can be defined analogously to the planar case. That is, a particle moves along geodesics inside  $Q$ , and reflects elastically upon hitting the boundary  $\partial Q$ . Suppose the previous reflection happens at  $\Gamma(s)$ . Let  $\theta$  be the angle measured from the (positive) tangent direction  $\dot{\Gamma}(s)$  to the post-reflection velocity of that particle. Then the *billiard map*  $F$  sends  $(s, \theta)$  to the next reflection  $(s_1, \theta_1)$  with  $\partial Q$ . The *phase space* of the billiard map  $F$  on  $Q$  is given by  $M = \Gamma \times (0, \pi)$ . Note that the 2-form  $\omega = \sin \theta \, ds \wedge d\theta$  is a symplectic form on  $M$ . Let  $\mu$  be the smooth probability measure on  $M$  with density  $d\mu = \frac{1}{2|\partial Q|} \sin \theta \, ds \, d\theta$ .

**Theorem 1.** *Let  $(S^2, g)$  be a convex sphere and  $Q \subset S^2$  be a strictly convex domain with  $C^r$  smooth boundary  $\Gamma = \partial Q$ . Then billiard map  $F : M \rightarrow M$  is a symplectic twist map. In particular,  $F$  preserves the measure  $\mu$ .*

It is well known that a twist map has periodic orbits of type  $(m, n)$  for all coprime pairs  $(m, n)$  [Bir27, Ban88]. We study some generic properties of the periodic points of dynamical billiards on a strictly convex domain  $Q$  on  $(S^2, g)$ . To this end, we identify the boundary  $\Gamma = \partial Q$  with the corresponding embedding function  $f : \mathbb{T} \rightarrow S^2$ . Let  $r \geq 2$  ( $r$  could be  $\infty$ ),  $\Upsilon^r(S^2, g)$  be the set of  $C^r$  smooth embeddings  $\Gamma \subset S^2$  such that the enclosed domains  $Q = Q(\Gamma)$  are strictly convex. Then  $\Upsilon^r(S^2, g)$  inherits a  $C^r$  topology from  $C^r(\mathbb{T}, S^2)$ .

**Theorem 2.** *There is a residual subset  $\mathcal{R}^r \subset \Upsilon^r(S^2, g)$ , such that for each  $\Gamma \in \mathcal{R}^r$ , the billiard map on  $\Gamma$  satisfies*

- (1) *each periodic point is either hyperbolic, or elliptic with irrational rotation number;*
- (2) *any two branches of invariant manifolds of hyperbolic periodic points either don't intersect, or they have some transverse intersections.*

Theorem 2 resembles the classical Kupka–Smale properties for dynamical billiards. The abstract Kupka–Smale property is proved by applying Thom Transversality Theorem, which requires the *richness* of local perturbations. On the other hand, dynamical billiards are known for the *lack* of local perturbations, since any perturbation of  $\Gamma$  results in a (semi)-global perturbation of the billiard map. See §4 for more details.

Given two hyperbolic periodic points  $p$  and  $q$ , these two points and their stable and unstable manifolds may be separated by some KAM-type invariant curves (which are persistent under perturbations). So the existence of heteroclinic intersections may not be generic. The following theorem answers positively the generic existence of homoclinic intersections.

**Theorem 3.** *There is a residual subset  $\mathcal{R}^r \subset \Upsilon^r(S^2, g)$ , such that for each  $\Gamma \in \mathcal{R}^r$ , there exist transverse homoclinic intersections for each hyperbolic periodic point of the billiard map  $F$  induced by  $\Gamma$ .*

The proof is based on Mather's characterization [Mat82] (developed by Franks and Le Calvez in [FL03]) of the *prime-end extension* of diffeomorphism on open surfaces. In his proof, Mather made an assumption that each *elliptic fixed point*, if exists, is Moser stable. This condition guarantees that there is no interaction between the hyperbolic and elliptic periodic points. To apply Mather's result, we have to study the elliptic periodic points first, although the hyperbolic ones are the subject we are interested in. This nonlinear stability is proved by one of Herman's result on Diophantine invariant curves.

Note that there are plenty of periodic points for twist maps, and hyperbolic periodic points exist generically. So the transverse homoclinic intersections in above theorem do exist generically.

**Corollary 1.** *There is an open and dense subset  $\mathcal{U}^r \subset \Upsilon^r(S^2, g)$ , such that for each  $\Gamma \in \mathcal{U}^r$ , the billiard map on  $\Gamma$  has positive topological entropy.*

Entropy is an important quantity indicating how chaotic a dynamical system is. The mechanism that a transverse homoclinic intersection generates chaos was first realized by Poincaré when he came across certain nonconvergent trigonometric series during his study of the  $n$ -body problem [Poin]. This mechanism was developed later by Birkhoff for the existence of infinitely many periodic points, and by Smale for his formulation of hyperbolic sets (horseshoe). Poincaré conjectured that for a generic  $f \in \text{Diff}_\mu^r(M)$ , and for every hyperbolic periodic point  $p$  of  $f$ ,

- (P1)  $W^s(p) \cap W^u(p) \setminus \{p\} \neq \emptyset$  (weaker version);
- (P2)  $W^s(p) \cap W^u(p)$  is dense in  $W^s(p) \cup W^u(p)$ .

This is the so called Poincaré's connecting problem. Poincaré also raised the closing problem about the denseness of periodic points, see [Pug67, PuRo83]. In the case  $r = 1$ , (P1) was proved by Takens in [Tak72]; (P2) was proved in [Tak72] on surfaces, and by Xia in [Xia96] on higher dimensional manifolds. For  $r \geq 2$ , most results about this connecting problem are on surfaces. Pixton proved in [Pix82] the property (P1) for planar surfaces, by extending Robinson's result [Rob73] on fixed points. For  $M = \mathbb{T}^2$ , (P1) was proved by Oliveira [Oli87]. For general surfaces, (P1) was proved by Oliveira [Oli00] for those with irreducible homological actions; and by Xia in [Xia06] for Hamiltonian diffeomorphisms. The proof of (P1) is still not complete for general surfaces, and there is almost no result on higher dimensions. The property (P2) is completely open even on surfaces. For planar convex billiards, (P1) was proved in [XZ14].

## 2. PRELIMINARIES

**2.1. Generating function of billiard map.** The billiard systems have an alternative definition using the generating function. More precisely, let  $(S^2, g)$  be a convex sphere, and  $Q \subset S^2$  be a strictly convex domain with  $C^r$  smooth boundary  $\Gamma = \partial Q$ . Let  $s \mapsto \Gamma(s)$  be the arc-length parameter. We will write  $s \in \Gamma$  by identifying  $s$  with  $\Gamma(s)$  if there is no confusion. For example, we set  $d_\Gamma(s_1, s_2) = d(\Gamma(s_1), \Gamma(s_2))$ . Let  $S(s_1, s_2) = -d_\Gamma(s_1, s_2)$ , and  $\partial_i S$  be the partial derivative of  $S$  with respect to  $s_i$ ,  $i = 1, 2$ . We extend the generating function to an arbitrary finite segment  $(s_m, \dots, s_n)$  with  $s_k \in \Gamma$ ,  $k = m, m+1, \dots, n$ , and define

the action functional  $W(s_m, \dots, s_n) = \sum_{k=m}^{n-1} S(s_k, s_{k+1})$  along the segment  $(s_k)_m^n$ . Such a segment is said to

be an orbit segment, if  $\partial_{s_k} W = \partial_2 S(s_{k-1}, s_k) + \partial_1 S(s_k, s_{k+1}) = 0$  for each  $k = m, \dots, n-1$ .

*Proof of Theorem 1.* Given two points  $s_1$  and  $s_2$ , let  $\gamma_1(t)$  be the geodesic from  $\gamma_1(0) = \Gamma(s_1)$  to  $\gamma_1(d) = \Gamma(s_2)$ , where  $d = d_\Gamma(s_1, s_2)$ . Let  $\theta_1$  be the angle from  $\dot{\Gamma}(s_1)$  to  $\dot{\gamma}_1(0)$ , and  $\theta_2$  be the angle from  $\dot{\Gamma}(s_2)$  to  $\dot{\gamma}_1(d)$ . At  $\Gamma(s_2)$ ,  $\gamma_1$  experiences an elastic reflection, and the new geodesic, say  $\gamma_2$ , starts from  $\gamma_2(0) = \Gamma(s_2)$ , such that the angle from  $\dot{\Gamma}(s_2)$  to  $\dot{\gamma}_2(0)$  equals  $\theta_2$ . One can check that

$$\partial_1 S(s_1, s_2) = \cos \theta_1, \quad \partial_2 S(s_1, s_2) = -\cos \theta_2. \quad (2.1)$$

Therefore,  $F(s_1, \theta_1) = (s_2, \theta_2)$  if and only if  $\partial_1 S(s_1, s_2) = \cos \theta_1$  and  $\partial_2 S(s_1, s_2) = -\cos \theta_2$ . Rewriting (2.1) in total differential form, we get  $dS = \cos \theta_1 ds_1 - \cos \theta_2 ds_2$ . Taking exterior differential and using  $d^2 S = 0$ ,

we get  $\sin \theta_2 ds_2 \wedge d\theta_2 = \sin \theta_1 ds_1 \wedge d\theta_1$ . Therefore, the 2-form  $\omega = \sin \theta ds \wedge d\theta$  is invariant under  $F$ , so is the probability measure  $d\mu = \frac{1}{2|\Gamma|} \sin \theta ds d\theta$  on  $M = \Gamma \times (0, \pi)$ .

To show that  $F$  is a twist map on  $M = \Gamma \times (0, \pi)$ , let's consider the image of  $M_s = \{s\} \times (0, \pi)$  under  $F$ . Let  $\gamma_\theta(t)$  be the geodesic starting from  $\Gamma(s)$  in the direction of  $\theta$ , and  $t_\theta > 0$  be the first moment that  $\gamma_\theta(t)$  hits  $\Gamma$ . The hitting position is exactly  $s_1(\theta) = \pi_1 \circ F(s, \theta)$ . Since  $Q$  is a strictly convex domain on  $S^2$ , the map  $s_1 : (0, \pi) \rightarrow \Gamma$  is monotonically increasing. Therefore,  $F$  is a symplectic twist map on  $M$ .  $\square$

**Corollary 2.** *Let  $\Gamma \in \Upsilon^r(S^2, g)$ , and  $F$  be the billiard map induced by  $\Gamma$ . Then for any coprime positive integers  $(p, q)$  with  $q \geq 2$ , there exists a periodic orbit  $\mathcal{O}_{p,q}$  of period  $q$  that goes around the table  $p$  times after one period.*

Such an orbit  $\mathcal{O}_{p,q}$  is called a Birkhoff periodic orbit of type  $(p, q)$ . See [Bir27, Ban88] for more details.

**2.2. Criterion of nondegenerate periodic orbits.** Let  $W(s_1, \dots, s_n) = \sum_{k=1}^n S(s_{k-1}, s_k)$  be the action on the space of the  $n$ -periodic configurations  $(s_k)$  in the sense that  $s_{n+k} = s_k$  for all  $k$ . Then  $x = (s, \theta) \in M$  is a periodic point with period  $n$  if and only if  $\partial_k W(s_1, \dots, s_n) = 0$  for each  $k = 1, \dots, n$ , where  $x_k = F^k x = (s_k, \theta_k)$  be the iterates of  $x$  under the billiard map. Given a critical  $n$ -periodic configuration  $(s_k)$ , we let  $D^2 W(s_1, \dots, s_n) = (\partial_{ij}^2 W)$  be the  $n \times n$  Hessian matrix of  $W$  at  $(s_1, \dots, s_n)$ .

Let  $D_x F^n$  be the tangent map at  $x$  (counted to its period), which is a  $2 \times 2$  matrix (in the coordinate system  $(s, \theta)$  on  $M$ ) with determinant 1 (since  $F$  preserves a smooth measure  $\mu$ ). Then  $x$  is said to be non-degenerate, if 1 is not an eigenvalue of  $D_x F^n$ . The later condition is equivalent to  $\text{Tr}(D_x F^n) \neq 2$  (since  $F$  preserves a smooth 2-form). Mackay and Meiss proved in [MM83] that the trace  $\text{Tr}(D_x F^n)$  is closely related to the Hessian  $D^2 W$  of  $W$  at its critical path  $(s_1, \dots, s_n)$ .

**Proposition 2.1.** *Let  $\{F^k x = (s_k, \theta_k)\}$  be a periodic orbit of period  $n$ ,  $W_2 = D^2 W(s_1, \dots, s_n)$  be the Hessian matrix of  $W$  at  $(s_1, \dots, s_n)$ . Then  $\text{Tr}(D_x F^n) - 2 = (-1)^n \cdot \det(W_2) \cdot \left( \prod_{i=1}^n S_{12}(s_{i-1}, s_i) \right)^{-1}$ .*

Note that  $\text{Tr}(D_x F^n) \neq 2$  if and only if  $\det W_2 \neq 0$ . So we have the following equivalent formulations:

- (1) a periodic orbit  $x = T^n x$  of the billiard map  $F$  is nondegenerate;
- (2) a critical cycle  $(s_1, \dots, s_n)$  of the action map  $W$  is nondegenerate.

Birkhoff made the following observation in [Bir27]. Let  $(s_1, \dots, s_n)$  be an  $n$ -periodic configuration at where  $W$  attains its minimum. Assume the corresponding periodic orbit  $x$  is nondegenerate. Then  $D^2 W(s_1, \dots, s_n)$  is positive definite, and  $\text{Tr}(D_x F^n) - 2 > 0$ . So the periodic point  $x$  corresponding to each minimizer turns out to be a hyperbolic periodic point.

**2.3. Curvature and focusing time of a tangent vector.** Now let's study some geometrical features of the tangent map of billiard map  $F : M \rightarrow M$  on the configuration space  $S^2$ . Let  $V \in T_x M$  be a tangent vector, and  $m(V)$  be the slope of  $V$  with respect to the coordinate  $(s, \theta)$ . By definition,  $V$  can be represented as  $V = \dot{c}(0)$ , where  $c : (-\epsilon, \epsilon) \rightarrow M$  is a smooth curve passing through  $c(0) = x$ . As in the planar case, each point  $x \in M$  represents a unit tangent vector at  $s = \pi_1(x)$ , and generates a geodesic segment  $\exp(t x)$  on  $Q$ . Then the above curve  $c$  generates a beam of billiard trajectories on  $Q$ , say  $\gamma_u(t) = \exp(t c(u))$ . A curve  $\rho : (-\epsilon, \epsilon) \rightarrow S^2$  with  $\rho(0) = \pi_1(x) = s$  and perpendicular to each  $\gamma_u$  at  $\rho(u)$  is called the wave-front corresponding to  $V \in T_x M$ . Let  $\mathcal{B}(V)$  be the geodesic curvature of  $\rho$  at  $p$ . The relation between  $\mathcal{B}(V)$  and the slope  $m(V)$  is given by  $m(V) = \mathcal{B}(V) \sin \theta + \kappa(s)$ , where  $V \in T_x M$  and  $s = \pi_1(x)$  is the projection to the first coordinate of  $x$ , see [ChMa06].

**Convention.** A wave-front has negative curvature if it is focusing, and has positive curvature if it is dispersing. Let  $\mathcal{B}(V) = \infty$  if  $p$  itself is a focusing point.

Any (infinitesimal) wave-front of billiard trajectories on  $Q$  focuses at some point forward and some point backward on  $S^2$  (not necessarily in  $Q$ ), say  $p_+$  and  $p_-$ . Let  $f(V) = d(p, p_+)$  be the forward focusing distance (time) of the wavefront related to  $V \in T_x M$ . In the case  $p$  itself is a focusing point, we set  $f(V) = 0$ .

Note that  $\mathcal{B}(V)$  and  $f(V)$  can be defined via normal Jacobi fields. That is, let  $J$  be the normal Jacobi field generated by the variation  $\gamma_u$  along  $\gamma_0$ . Then  $\mathcal{B}(V) = \frac{J'(0)}{J(0)}$ , and  $f(V) = \min\{t \geq 0 : J(t) = 0\}$ , and  $J(t)$

satisfies the Jacobi equation  $J'' + K_g \cdot J = 0$ , where  $K_g$  is the Gaussian curvature of  $(S^2, g)$ . So the relation between  $\mathcal{B}(V)$  and  $f(V)$  is given by the solution of the Jacobi equation. For example, if  $\mathcal{B}(V) = 0$  then the wavefront focuses at two *focal points* along the geodesic  $\gamma_x$  (one forward focal point, and one backward focal point), and these two focal points are conjugate along  $\gamma_x$ .

Now let  $x = (s, \theta) \in M$ ,  $Fx = (s_1, \theta_1)$ ,  $V \in T_x M$ ,  $V_1 = DF(V) \in T_{x_1} M$ , and  $\rho$  be a wavefront related to  $V$ . Let  $\mathcal{B}_t(V)$  and  $f_t(V)$  be the curvature and forward focusing time of the wavefront during the free flight time  $0 < t < d_1 = d_\Gamma(s, s_1)$ ,  $\mathcal{B}^\pm(V_1)$  and  $f^\pm(V_1)$  be the curvature and focusing time right before/after the collision  $t \rightarrow d_1 \pm 0$ . Then we have

- (1).  $\mathcal{B}_t(V) = \frac{J'(t)}{J(t)}$ , where  $J(t)$  is the solution of Jacobi equation;
- (2).  $\mathcal{B}^+(V_1) = \mathcal{B}^-(V_1) - \frac{2\kappa(s_1)}{\sin \theta_1}$ , where  $\kappa(s_1) > 0$  is the curvature at  $\Gamma(s_1)$ .

Item (2) is the so called Mirror Formula for geometrical optics on surfaces, see [Vet84]. Note that  $f_t(V) = f(V) - t$  when  $t \leq f(V)$ . If  $f(V) < d_\Gamma(s, s_1)$ , then the wavefront focuses between two consecutive reflections,  $\mathcal{B}_t(V)$  jumps from  $-\infty$  to  $+\infty$ , and  $f_t(V)$  jumps from 0 to the next focusing time.

**Example.** In the case that  $g = g_0$  is the round metric on  $S^2$ , the quantities  $\mathcal{B}(V)$ ,  $f(V) = d(p, p_+)$  and  $\hat{f}(V) = d(p, p_-)$  are related by the following formula:

$$f(V) + \hat{f}(V) = \pi, \quad \mathcal{B}(V) = -\cot f(V) = \cot \hat{f}(V). \quad (2.2)$$

Let  $\mathcal{B}(V) = \cot \alpha_0$ . Then  $\mathcal{B}_t(V) = \cot(\alpha_0 + t)$  for all  $0 \leq t < d(s, s_1)$ .

*Proof of (2.2).* Let's consider the circles  $L_\alpha$  of latitude on  $S^2$  surrounding the north pole, where  $\alpha$  is the angle of the circle with the positive  $z$ -axis. Then the radius of  $L_\alpha$  is  $r(\alpha) = \sin \alpha$ , and the geodesic curvature is  $\kappa(\alpha) = \sqrt{1/r^2 - 1} = \cot \alpha$ . Then the results follow from the observation that  $d(p, p_+) = \alpha$  and  $d(p, p_-) = \pi - \alpha$  (and the convention on the choices of signs of the curvature).  $\square$

**2.4. Some generic properties of periodic orbits.** Let  $(S^2, g)$  be a convex sphere,  $Q \subset S^2$  be a strictly convex domain, and  $F : M \rightarrow M$  be the induced billiard map on  $Q$ , where  $M = \Gamma \times (0, \pi)$ . Let  $p$  be a periodic point and  $\mathcal{O}(p)$  be the orbit of  $p$ . There are some special features for the periodic orbits on the billiard map  $Q$  (see also [Sto87]):

- (1)  $|\mathcal{O}(p)| \neq |\pi_1(\mathcal{O}(p))|$ : the orbit passes some reflection point more than once during a minimal period.
- (2)  $|\pi_1(\mathcal{O}(p) \cup \mathcal{O}(q))| \neq |\pi_1(\mathcal{O}(p))| + |\pi_1(\mathcal{O}(q))|$ : two different periodic orbits have some common reflection points.

Take the round table on standard sphere for example: on each point  $s \in \Gamma$ , there exist periodic orbits of type  $(m, n)$  for all  $(m, n)$ . This happens even among the orbits with the same period: the  $(1, 5)$ -orbit (pentagon) and the  $(2, 5)$ -orbit (pentagram).

Before giving the precise definition, we need to distinguish the following two cases: symmetric and non-symmetric orbits. A periodic orbit  $\mathcal{O}(p)$  is said to be *symmetric*, if  $\theta_k = \pi/2$  for some  $k$ . Along such an orbit, the period  $n = 2m$  is an even number, the right angle reflections happen exactly twice, and the orbit travels back and forth between these two reflection points. See [Sto87]. A periodic orbit is said to be *nonsymmetric*, if it is not symmetric.

**Definition 2.** If a periodic orbit  $\mathcal{O}(p)$  is nonsymmetric, then the defect of  $p$  is defined by the difference  $d(p) = |\mathcal{O}(p)| - |\pi_1(\mathcal{O}(p))|$ . If  $\mathcal{O}(p)$  is symmetric, then the defect of  $p$  is defined by  $d(p) = \frac{1}{2}|\mathcal{O}(p)| + 1 - |\pi_1(\mathcal{O}(p))|$ .

See Fig. 1 for a schematic sketch of (planar) periodic orbits with positive defect: (A) for nonsymmetric case, and (B) for symmetric case.

**Proposition 2.2.** Let  $P_n(\Gamma)$  be the set of points fixed by  $F^n$ . There is a residual subset  $\mathcal{S}_n \subset \Upsilon^r(S^2, g)$ , such that the following hold for the billiard map of each  $\Gamma \in \mathcal{S}_n$ ,

- (1) every periodic orbit in  $P_n(\Gamma)$  has zero defect;
- (2) two different periodic orbits in  $P_n(\Gamma)$  don't pass through any common reflection point.

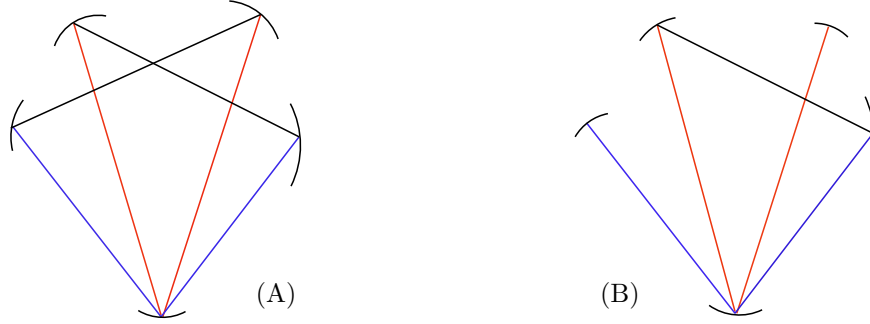


FIGURE 1. Periodic orbits with positive defects. (A): nonsymmetric case; (B): symmetric case.

Note that the periodic orbits of period 2 always have zero defect. So  $\mathcal{S}_2 = \Upsilon^r(S^2, g)$ .

For billiards in the Euclidean domain, Proposition 2.2 have been proved by Stojanov [Sto87]. Note that the following two statements are equivalent for a given  $\Gamma$ :

- every periodic orbit has zero defect;
- any closed path  $s_0, s_1, \dots, s_n = s_0$  with positive defect is not a real orbit.

Then Proposition 2.2 is proved by showing that the second statement holds generically. The proof for billiards on  $S^2$  follows from the same idea, and is sketched in the Appendix.

**Remark 1.** Let  $\mathcal{S} = \bigcap_{n \geq 1} \mathcal{S}_n$ , which contains a residual subset of  $\Upsilon^r(S^2, g)$ . Then for each  $\Gamma \in \mathcal{S}$ ,

- (a) every periodic orbit of  $F$  has zero defect;
- (b) two different periodic orbits of  $F$  don't pass any common reflection point.

One would expect that  $\mathcal{S}_n$  could be open and dense, not just residual. However, this may not be true for general domains. In next section we will prove that the properties (a) and (b) do hold on an open and dense subset of the convex domains in  $\Upsilon^r(S^2, g)$ .

**Remark 2.** The following properties are obtained in [PeSt87a, PeSt87b, PeSt88] for billiard systems on a generic connected domain in  $\mathbb{R}^d$ :

- (c) the set of points fixed by  $F^n$  is finite;
- (d) the eigenvalue of each periodic point fixed by  $F^n$  is not in  $\mathcal{A}$ ,

where  $\mathcal{A}$  is any countable subset of  $\mathbb{R}$  given in advance. The 2D version has been obtained by Lazutkin [Laz81]. We will prove that these properties hold on an open and dense subset of convex billiards, and the sets of points fixed by  $F^n$  actually vary continuously. This continuity plays a key role in the study of homoclinic and heteroclinic intersections.

### 3. PERTURBATIONS OF PERIODIC POINTS OF BILLIARD SYSTEMS

There are various types of perturbation techniques in the study of dynamical systems. One of the widely used technique is *Franks' Lemma*, which allows us to manipulate the derivatives along a periodic orbit. The perturbations for billiard dynamics are very limited, since one can't perturb the billiard map  $F$  directly, while the perturbation of the underlining table changes the dynamics (semi)-globally. See Visscher's thesis [Vis12] for several results on Franks's lemma in geometric contexts (geodesics flows and billiards). In [DOP07] the effect of the perturbation of a planar billiard system is computed explicitly via a step by step induction. It is difficult to generalize their approach to dynamical billiards on surfaces with non-constant curvatures. In this section we present another proof, which uses the geometric features of the tangent vectors of the phase space  $M$  on the configuration space  $S^2$ .

We first give some basic definitions. Let  $p$  be a periodic point of  $F$  of period  $n$ ,  $D_p F^n : T_p M \rightarrow T_p M$  be the tangent map, which can be viewed as a matrix in  $\text{SL}(2, \mathbb{R})$ . Let  $\lambda_p$  be an eigenvalue of  $D_p F^n$ . Then

$p$  is said to be *hyperbolic* if  $|\lambda_p| \neq 1$ , be *parabolic* if  $\lambda_p = \pm 1$ , and be *elliptic* if otherwise. Alternatively, a periodic point  $p$  is said to be *degenerate* if  $\lambda_p = 1$ , and be *nondegenerate* if it is not degenerate.

Let  $\tau(p)$  be the trace of  $D_p F^n$ . Then we have the following equivalent definition:  $p$  is said to be hyperbolic if  $|\tau(p)| > 2$ , be parabolic if  $|\tau(p)| = 2$ , be elliptic if  $|\tau(p)| < 2$ , be degenerate if  $\tau(p) = 2$ , and be nondegenerate if  $\tau(p) \neq 2$ . All nondegenerate periodic points persist under small perturbations.

**3.1. Useful perturbations of billiard systems.** The following perturbations are widely used in the study of generic properties of billiards.

**Definition 3.** Let  $s_0 \in \Gamma$ , and  $I \subset \Gamma$  be a neighborhood of  $s_0$ . Then a *normal perturbation*  $\Gamma_\epsilon$  of  $\Gamma$  at  $s_0$  supported on  $I$  is a convex curve on  $S^2$  that satisfies  $\Gamma_\epsilon(s) = \Gamma(s)$  for  $s_0$  and for  $s \notin I$ ,  $\dot{\Gamma}_\epsilon(s_0) = \dot{\Gamma}(s_0)$ , while the curvature changes to  $\kappa_\epsilon(s_0) = \kappa(s_0) + \epsilon$ .

In fact, the normal perturbations are essentially the only type of perturbations that preserve the orbit  $\mathcal{O}(p)$ , in the meanwhile, change the derivatives of  $DF^n$  at  $p$ . However, a degenerate periodic point may be *robustly degenerate* under normal perturbations.

**Example.** Let  $\gamma$  be a geodesic starting at a point  $\mathbf{p} \in S^2$ , and  $\mathbf{q}$  be a conjugate point of  $\mathbf{p}$  along  $\gamma$ . Let  $Q \subset S^2$  be a convex domain containing the geodesic segment  $\gamma$  from  $\mathbf{p}$  to  $\mathbf{q}$  as a diameter. Then there is a periodic orbit of period 2 traveling along  $\gamma$  back and forth. Let  $p = (\mathbf{p}, \pi/2)$  be the corresponding point on the phase space  $M$ . Then the wavefront leaving  $\mathbf{p}$  as a focusing point will bounce back and forth between these two reflection points  $\mathbf{p}$  and  $\mathbf{q}$ , and focus at each reflection. If  $p$  is a degenerate periodic point for  $F$ , then the degeneracy of  $p$  persists under normal perturbations.

*Proof.* Our proof actually works for any period. This general formulation will be used later. Let  $p$  be a periodic point such that there is no multiple reflections at  $s_0 = \pi_1(p)$ ,  $\Gamma_\epsilon$  be a normal perturbation of  $\Gamma$  at  $s_0$ . Then for each  $V \in T_p M$ , the total effect of  $D_p F_\epsilon^n$  on  $V$  is a shift of the curvature of the returning wave-front of  $D_p F^n(V)$ :  $\mathcal{B}(D_p F_\epsilon^n(V)) = \mathcal{B}(D_p F^n(V)) - \frac{2\epsilon}{\sin \theta}$ , and a shift of the slope  $m(D_p F_\epsilon^n(V)) = m(D_p F^n(V)) + \epsilon$ . Therefore,  $D_p F_\epsilon^n = \pm \begin{bmatrix} 1 & 0 \\ \epsilon & 1 \end{bmatrix} \circ D_p F^n$ . Then the sign is positive, since  $\Gamma_\epsilon$  is a small perturbation of  $\Gamma$ .

In the setting of the above example, we denote  $D_p F^2 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ . Then  $b = 0$  since the line  $\langle \partial_\theta \rangle$  is invariant, and  $a = d = 1$ , since  $a + d = 2$  (degeneracy assumption) and  $ad = 1$  (symplectic property). Therefore,  $D_p F^2 = \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}$  and  $D_p F_\epsilon^2 = \begin{bmatrix} 1 & 0 \\ \epsilon + c & 1 \end{bmatrix}$ . This implies that  $p$  is degenerate for any normal perturbation.  $\square$

This type of persistence of degeneracy of periodic orbits (with higher periods) may happen for the convex billiards on  $S^2$  and for planar billiards. To overcome this difficulty, we need to consider another type of perturbations, which shift the base point  $s_0$  along the normal direction at  $\Gamma(s_0)$ . It is very likely that, after the shifting perturbation, the orbit passing through  $p$  is not even closed. Luckily for us, such a shift is only needed when the reflection at  $p$  is the right angle, and there is no multiple reflections at  $s_0 = \pi_1(p)$  in the period of  $p$ . In (and only in) this case, the periodic orbit  $\mathcal{O}(p)$  stays the same after the shift of  $\Gamma$  along the normal direction at  $\Gamma(s_0)$ .

**3.2. Perturbations of periodic points.** Let  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \text{SL}(2, \mathbb{R})$ , and  $A_\epsilon = \begin{bmatrix} 1 & 0 \\ \epsilon & 1 \end{bmatrix} \circ A$ . Then  $\text{Tr}(A) = a + d$  and  $\text{Tr}(A_\epsilon) = a + d + \epsilon b$ . Given a periodic point  $p$  of period  $n$ , we let  $D_p F^n = \begin{bmatrix} a_p & b_p \\ c_p & d_p \end{bmatrix}$ , and denote  $\tau(p) := \text{Tr}(D_p F^n) = a_p + d_p$ .

Note that the dynamics near a hyperbolic periodic point is topologically conjugate to the linearized map  $D_p F^n$  (by Hartman–Grobman Theorem) and is well understood. However, the dynamics surrounding the degenerate and elliptic ones are quite complicated, very sensitive to the arithmetic properties of the linearization of  $F^n$  at  $p$ , and depend on the nonlinear part of  $F$ .

**Proposition 3.1.** *Let  $\Gamma \in \Upsilon^r(S^2, g)$ , and Let  $p$  be a periodic point of the billiard map  $F$  with zero defect. Suppose  $p$  is not hyperbolic. Then there is a small perturbation  $\Gamma_\epsilon$  of  $\Gamma$  such that the trace  $\tau_\epsilon(p) \neq \tau(p)$ .*

In other words, we have the following qualitative descriptions:

- (1) if  $p$  is degenerate, then after the perturbation, it is either hyperbolic or elliptic;
- (2) if  $p$  is elliptic, then the rotation number of  $p$  can be shifted continuously under the perturbation.

*Proof.* Let  $p$  be a periodic point with period  $n$ . Let  $\Gamma_\epsilon$  be a normal perturbation of  $\Gamma$  at  $s_0 = \pi_1(p)$ . Then we have

$$\tau_\epsilon(p) = \text{Tr}(D_p F_\epsilon^n) = a_p + d_p + \epsilon b_p.$$

If  $b_p \neq 0$ , then we are done. In the following we assume  $b_p = 0$ .

If  $b_p = 0$ , then we have  $a_p \cdot d_p = 1$ , which implies  $|a_p + d_p| \geq 2$ . Note that  $p$  is assumed to be non-hyperbolic. So we actually have  $|a_p + d_p| = 2$ , and  $D_p F^n = \pm \begin{bmatrix} 1 & 0 \\ c_p & 1 \end{bmatrix}$ . Then the line  $\langle \partial_\theta \rangle_p$  is fixed by  $D_p F^n$ . Equivalently, the corresponding wavefront  $\rho_p$  focuses at  $s_0 = \pi_1(p)$ , and will focus at  $s_0$  again when it returns after one period. So we only need to show that a small perturbation can destroy the last property (for some point on the orbit) of  $p$ .

**Case 1.** The orbit of  $p$  is not symmetric. Then the zero defect property implies that there is no multiple reflection along the orbit  $\mathcal{O}(p)$ .

**Case 1a.**  $c_p \neq 0$ . In this case,  $\langle \partial_\theta \rangle_p$  is the only line fixed by  $D_p F^n$ , and  $\rho_p$  is the only invariant wavefront at  $p$  and along the whole orbit of  $p$ . Clearly this wavefront does not focus at  $s_1 = \pi_1(Fp)$ , since there is no conjugate point on  $\Gamma$ . Therefore  $\langle \partial_\theta \rangle_{Fp}$  is not fixed by  $D_{Fp} F^n$ , since the wavefront corresponding to  $\langle \partial_\theta \rangle_{Fp}$  focuses at  $s_1$  (hence is not invariant). This implies  $b_{Fp} \neq 0$ , and a normal perturbation  $\Gamma_\epsilon$  of  $\Gamma$  is performed at  $s_1$ . Then  $\tau_\epsilon(Fp) = \tau(Fp) + \epsilon \cdot b_{Fp}$ , and the proposition follows since  $\tau(Fp) = \tau(p)$  and  $\tau_\epsilon(Fp) = \tau_\epsilon(p)$ .

**Case 1b.**  $c_p = 0$ . In this case  $D_p F^n = \pm I_2$ . We first make a normal perturbation at  $s_0$ , and get  $D_p F_\epsilon^n \neq \pm I_2$ . Then we do another perturbation given as in Case 1a.

**Case 2.** Now we assume that the orbit of  $p$  is symmetric. The difficulty in this case is that there may exist multiple reflections. Without loss of generality we assume  $n > 2$ . In this case, there are exactly two simple reflections among the orbit  $\mathcal{O}(p)$ , and all other reflections happens twice (forward and backward). Moreover,  $\pi_2(F^k p) = \pi/2$  at those two simple reflections. This time, it is possible that a wavefront focuses exactly at these two ends (say  $s_0$  is one of them). If this happens, we make an extra shift of  $\Gamma$  along the normal direction at  $s_0$ , so that the focusing wavefront is not invariant any more. Then one also gets  $b_p \neq 0$  and  $\tau_\epsilon(p) = \tau(p) + \epsilon \cdot b_p$ . This completes the proof.  $\square$

#### 4. KUPKA-SMALE PROPERTIES FOR CONVEX BILLIARDS

Let  $Q \subset S^2$  be a strictly convex domain with  $C^r$  smooth boundary  $\Gamma$ ,  $M = \Gamma \times (0, \pi)$  be the phase space of the billiard map  $F$  induced on  $Q$ . For each  $n \geq 2$ , let  $P_n(\Gamma) \subset M$  be the set of points fixed by  $F^n$ . In the following we will show that there is an open and dense subset  $\mathcal{U}_n$  such that for each  $\Gamma \in \mathcal{U}_n$ ,  $P_n(\Gamma)$  is finite and depends continuously on  $\Gamma$ .

**Remark 3.** The geodesic on Riemannian manifolds are time-reversal invariant (this may not be true on general Finsler manifolds). Similarly, the billiard dynamics on a convex table  $Q \subset S^2$  is time-reversal invariant. More precisely, let  $\Theta : M \rightarrow M$ ,  $(s, \theta) \mapsto (s, \pi - \theta)$  be the time-reversal map. Then  $F \circ \Theta = \Theta \circ F^{-1}$ . So if  $x$  is a periodic point of  $F$ , so is  $\Theta(x)$  (the two orbits are distinct if  $\pi/2 \notin \pi_2(\mathcal{O}(x))$ ). Moreover, these two have the same period and same stability property. We only need to consider one of them when making perturbations.

For an abstract diffeomorphism  $f$  on  $M$ , the periodic set  $P_n(f)$  may not be compact, since the phase space  $M$  is not compact<sup>1</sup>. We will show that this can't happen due to a special feature that is only true for convex billiards [DOP07].

<sup>1</sup> We don't want to take the closure of  $M$ , since every point on the added boundary  $\partial M = \Gamma \times \{0, \pi\}$  is fixed and hence degenerate.



**4.1. Compactness of  $P_n(\Gamma)$  for all  $\Gamma$ .** Let  $\Gamma \in \Upsilon^r(S^2, g)$  and  $Q \subset S^2$  be the strictly convex domain enclosed by  $\Gamma$ . Note that  $\kappa(s) > 0$  and  $\int_Q K_g d\sigma + \int_\Gamma \kappa(s) ds = 2\pi$  by the Gauss–Bonnet Theorem.

Let  $F$  be the billiard map on  $Q$ . Note that  $F$  has no fixed point in  $M = \Gamma \times (0, \pi)$ , and  $P_2(\Gamma) \subset \Gamma \times \{\pi/2\}$  is always closed and compact. In the following we let  $n \geq 3$ .

Denote  $\delta^* = \min\{\pi_2 \circ F(s, \pi/2) : s \in \Gamma\}$ . It is easy to see that  $0 < \delta^* \leq \pi/2$ . Then for each  $n \geq 3$ , let  $\delta_n = \frac{1}{2n} \int_\Gamma \kappa(s) ds > 0$ , and  $M_n = \Gamma \times [\delta_n^*, \pi - \delta_n^*]$  be the central annulus of the phase space, where  $\delta_n^* = \min\{\delta^*, \delta_n\}$ .

**Lemma 4.1.** *Let  $n \geq 3$ . Then  $\mathcal{O}(x) \cap M_n \neq \emptyset$  for each  $x \in P_n(\Gamma)$ .*

*Proof.* Note that  $M_n$  is time-reversal invariant, and increases with  $n$ . So we only need to consider those points  $x \in P_n(\Gamma)$  with minimal period  $n$ . Let  $F^i x = (s_i, \theta_i)$  for each  $1 \leq i \leq n$ . Note that  $\theta_i \neq \pi/2$  for some  $i$ . By time-reversal invariance of  $F$ , we can assume that  $\theta_i < \pi/2$  for some  $i$ .

There are two cases: 1)  $\theta_{i-1} \geq \pi/2 > \theta_i$  for some  $i$ , or 2)  $\theta_i < \pi/2$  for all  $i$ .

Case 1). By the definition of  $\delta^*$ , one has that  $\theta_i \geq \delta^*$  whenever  $\theta_{i-1} \geq \pi/2$ . So we have  $\delta^* \leq \theta_i < \pi/2$  and hence  $F^i x \in M_n$ .

Case 2). Now we assume  $\theta_i < \pi/2$  for all  $i$ . Let  $\Delta_x$  be the polygon on  $Q$  traced out by the orbit of  $x$ .

(2a).  $\Delta_x$  is geodesically convex. Since all the edges of  $\Delta_x$  are parts of geodesics, we have  $2 \sum_{i=0}^{n-1} \theta_i + \int_{\Delta_x} K_g d\sigma = 2\pi$  (compare with  $\int_\Gamma \kappa(s) ds + \int_Q K_g d\sigma = 2\pi$ ). Clearly  $\Delta_x \subset Q$  since  $Q$  is strictly convex. So  $2 \sum_{i=0}^{n-1} \theta_i \geq \int_\Gamma \kappa(s) ds$ , which implies  $\max_{0 \leq i < n} \theta_i \geq \delta_n$ . Therefore,  $F^i x \in M_n$  for some  $i$ .

(2b).  $\Delta_x$  may not be convex. Let  $\Delta_x^*$  be the geodesic convex hull of  $\Delta_x$ . Clearly  $\Delta_x \subset \Delta_x^*$ . Note that the orbit  $\mathcal{O}(x)$  may have positive defect, and the boundary of  $\Delta_x^*$  may not generate real orbit. Let  $s_i$ ,  $1 \leq i \leq v$  be vertices of  $\Delta_x^*$ , and  $\alpha_i, \beta_i$  be the incoming and leaving angle at  $s_i$  along the boundary of  $\Delta_x^*$ . Then we have  $\sum_{i=1}^d (\alpha_i + \beta_i) + \int_{\Delta_x^*} K_g d\sigma = 2\pi$ , and hence  $\sum_{i=1}^d (\alpha_i + \beta_i) \geq \int_\Gamma \kappa(s) ds$ . Note that for each  $F^j x$  with  $\pi_1(F^j x) = s_i$ , we have  $\theta_j \geq \max\{\alpha_i, \beta_i\}$ . Therefore,

$$\max_{0 \leq j < n} \theta_j \geq \max_{1 \leq i \leq d} \max\{\alpha_i, \beta_i\} \geq \delta_n.$$

So  $F^j x \in M_n$  for some  $j$ .

Putting them together, we get that for any periodic point  $x \in P_n(\Gamma)$ , there exists an  $i$  with  $\theta_i \in [\delta_n, \pi - \delta_n]$  and hence  $F^i x \in M_n$ . This completes the proof.  $\square$

**Corollary 3.** *For each  $n \geq 2$ , the set  $P_n(\Gamma)$  is compact.*

*Proof.* It suffices to note that  $P_n(\Gamma) \cap M_n$  is compact, and  $P_n(\Gamma) = \bigcup_{k=0}^{n-1} F^k(P_n(\Gamma) \cap M_n)$ .  $\square$

**4.2. Finiteness of  $P_n(\Gamma)$  for most  $\Gamma$ .** A periodic point  $x$  is said to be *non-degenerate*, if 1 is not an eigenvalue of  $D_x F^{m(x)} : T_x M \rightarrow T_x M$ , where  $m(x)$  be the minimal period of  $x$ . The minimal period of a periodic point  $x \in P_n(\Gamma)$  satisfies  $m(x)|n$ , and may be strictly less than  $n$ . Then  $x$  is said to be *non-degenerate under  $F^n$* , if 1 is not an eigenvalue of  $D_x F^n : T_x M \rightarrow T_x M$ .

Let  $\mathcal{U}_n \subset \Upsilon^r(S^2, g)$  be the set of strictly convex domains  $\Gamma \in \Upsilon^r(S^2, g)$  such that every periodic point  $x \in P_n(\Gamma)$  is non-degenerate under  $F^n$ . Now we state the first main result of this section.

**Proposition 4.2.** *The set  $\mathcal{U}_n$  is an open and dense subset of  $\Upsilon^r(S^2, g)$ .*

The proof of Proposition 4.2 is quite long. We first prove some basic properties of the billiard maps for  $\Gamma \in \mathcal{U}_n$ .

**Lemma 4.3.** *The set  $P_n(\Gamma)$  is a finite set for each  $\Gamma \in \mathcal{U}_n$ , and the map  $\Gamma \mapsto P_n(\Gamma)$  is continuous on  $\mathcal{U}_n$ .*

*Proof.* Note that some points in  $P_n(\Gamma)$  may be fixed by  $m|n$ . Suppose on the contrary that there do exist infinitely many such orbits, say  $\mathcal{O}_k \subset P_n(\Gamma)$  for each  $k \geq 1$ . Pick  $x_k \in \mathcal{O}_k \cap M_n$  and assume that  $x_k \rightarrow x$ . Note that  $x \in M_n$  and  $F^n x = \lim F^n x_k = \lim x_k = x$ . Therefore, the point  $x$  is nondegenerate under  $F^n$ , and we can find a neighborhood  $U \subset M$  of  $x$ , such that the continuation  $x$  is the only point fixed by  $F^n$  in  $U$ . This contradicts the choice of  $x$  as a limiting point of  $s_k$ .

For the continuity, let's consider the map  $P_n : \Gamma \in \Upsilon^r(S^2, g) \mapsto P_n(\Gamma)$ . Let  $\Gamma \in \mathcal{U}_n$ . Then

- upper semicontinuous: pick a sequence  $\Gamma_k \rightarrow \Gamma$ , and a sequence  $x_k \in P_n(\Gamma_k) \rightarrow x$ . Then  $F^n x = x$  and hence  $x \in P_n(\Gamma)$ . Therefore  $\limsup_{k \rightarrow \infty} P_n(\Gamma_k) \subset P_n(\Gamma)$ ;
- lower semicontinuous: every point  $x \in P_n(\Gamma)$  is nondegenerate under  $F^n$ , and nondegenerate periodic point persists for nearby systems. Therefore  $\liminf_{\Gamma' \rightarrow \Gamma} P_n(\Gamma') \supset P_n(\Gamma)$ .

Combining these two terms, we have that  $P_n(\Gamma)$  varies continuously on  $\mathcal{U}_n$ .  $\square$

#### 4.3. Proof of Proposition 4.2.

We first prove the openness of  $\mathcal{U}_n$ .

*Proof of Proposition 4.2: the openness.* We prove by contradiction. Suppose there exist  $\Gamma \in \mathcal{U}_n$  and a sequence of  $\Gamma_k \in \Upsilon^r(S^2, g) \setminus \mathcal{U}_n$  with  $\Gamma_k \rightarrow \Gamma$ . Since  $\Gamma_k \notin \mathcal{U}_n$ , there exists some degenerate periodic point  $x_k \in P_n(\Gamma_k)$  with minimal period  $m_k|n$ . By Lemma 4.1, one can assume  $x_k \in M_n$ . Passing to a subsequence if necessary, we assume  $m_k \equiv m$ ,  $x_k \rightarrow x \in M_n$ . Then we get  $F^n x = \lim_{k \rightarrow \infty} F_k^n x_k = \lim_{k \rightarrow \infty} x_k = x$ . The point  $x$  is nondegenerate under  $F^n$ , and there exists a neighborhood  $U \subset M$  of  $x$ , such that the continuation  $x(\Gamma_k)$  is nondegenerate under  $F_k^n$ , and is the only point fixed by  $F_k^n$  in  $U$ . This contradicts the fact that  $x_k$  is degenerate under  $F_k^n$  and will enter  $U$  eventually.  $\square$

Note that no bifurcation of periodic points is allowed in  $\mathcal{U}_n$ . So we have the following corollary.

**Corollary 4.** *The cardinal map  $\Gamma \in \mathcal{U}_n \rightarrow |P_n(\Gamma)|$  is locally constant.*

Now we are left with the proof of the denseness of  $\mathcal{U}_n$ . We first give a direct proof for  $n = 2$  to illustrate the idea of the proof. The proof for the general case is given after that.

*Proof of Proposition 4.2: the denseness of  $\mathcal{U}_2$ .* Let  $\Gamma \in \Upsilon^r(S^2, g)$  be parameterized by  $\mathbb{T} \rightarrow S^2$ . Given  $\epsilon > 0$ , let  $I_i = (s_i - \epsilon, s_i + \epsilon)$ , and  $I_i^r = (s_i - \epsilon^r, s_i + \epsilon^r)$  be the core of  $I_i$  such that  $\bigcup_{i=0}^{n_2} I_i^r$  covers  $\mathbb{T}$  for some  $n_2 \geq 1$ .

We pick  $\epsilon$  small enough such that the normal geodesics (that is,  $\theta = \pi/2$ ) hit each arc  $I_i$  no more than once. Then we cover the central line  $M_{\pi/2} = \Gamma \times \{\pi/2\} \subset M$  by much smaller disks  $\{B_j = B(x_j, \delta_j) : 1 \leq j \leq m_2\}$  such that for each  $k = 0, 1$ ,  $\pi_1(F^k B(x_j, \delta_j)) \subset I_{i(k)}^r$  for some  $i = i(k)$ .

Given  $(s, \alpha)$  and  $i = 1 \dots, m_2$ , let  $\Gamma_i(s, \alpha)$  be a perturbation of  $\Gamma$  supported on  $I_i$  that shifts  $I_i^r$   $s$  distance along the geodesic passing through  $s_i$  in the direction of  $\theta = \pi/4$ , then rotates the tangent direction of the resulting curve by an angle  $\alpha$ . The exact number in  $\theta = \pi/4$  is not important, as long as  $\theta \neq \pi/2$ . Since  $\Upsilon^r(S^2, g)$  is open, there exists an open disk  $D_i \subset \mathbb{R}^2$  of  $(s, \alpha) = (0, 0)$ , such that  $\Gamma_i(s, \alpha) \in \Upsilon^r(S^2, g)$  is  $(C^r, \epsilon)$ -close to  $\Gamma$ . Let  $F_{i,s,\alpha}$  be the billiard map induced by  $\Gamma_i(s, \alpha)$ . This gives rise to a map  $\zeta_i : (s, \alpha) \in D_i \rightarrow F_{i,s,\alpha}$ , and an evaluation map  $\zeta_i^{\text{ev}} : D_i \times M \rightarrow M$ ,  $(s, \alpha, x) \mapsto F_{i,s,\alpha}(x)$ .

For each  $x \in P_2(\Gamma)$ , there exists  $i$  such that  $\pi_1(x) \in I_i^r$ , while  $\pi_1(Fx) \notin I_i$  (by the choice of  $I_i$ ). In particular,  $F_{i,s,\alpha} \equiv F$  on the set of points not based on  $I_i$ . Then  $U = \{F_{i,s,\alpha}x : (s, \alpha) \in D_i\}$  is an open neighborhood of  $Fx$ , and  $\{F_{i,s,\alpha}^2x : (s, \alpha) \in D_i\} = FU$  is still an open neighborhood of  $F^2x$ . Therefore the evaluation map  $\zeta_i^{\text{ev}} : D_i \times M \rightarrow M \times M$ ,  $(s, \alpha, x) \mapsto (x, F_{i,s,\alpha}^2(x))$  is transverse to  $\Delta \subset M \times M$  at  $(0, 0) \times B_i$ . Putting all these  $(D_i, \zeta_i)$  together, we get an evaluation map

$$\zeta^{\text{ev}} : D_1 \times \dots \times D_{m_2} \times M \rightarrow M \times M, ((s_j, \alpha_j)_{j=1}^{m_2}, x) \mapsto (x, F_{(s_j, \alpha_j)}^2(x)),$$

which is transverse to  $\Delta$  along  $0^{2m_2} \times \bigcup_j B(x_j, \delta_j) \supset \mathbf{0} \times M_{\pi/2}$ . In particular, there exists an open neighborhood  $D \subset D_1 \times \dots \times D_{m_2}$  such that the evaluation map is transverse to  $\Delta$  along  $D \times M_{\pi/2}$ . Then by the Parametric Transversality Theorem (see [Rob95]), there is a residual set of parameters  $(s_j, \alpha_j)_{j=1}^{m_2}$  such

that the graphs of  $F_{(s_j, \alpha_j)}^2$  are transverse to  $\Delta$ . In particular,  $\Gamma_{(s_j, \alpha_j)} \in \mathcal{U}_2$  and they approximate  $\Gamma$ . This shows that  $\mathcal{U}_2$  is dense in  $\Upsilon^r(S^2, g)$ .  $\square$

We prove the denseness of  $\mathcal{U}_n$ ,  $n \geq 3$  by strong induction. Suppose that we have proven the existence of an open and dense subset  $\mathcal{U}_k$  for each  $2 \leq k < n$  with  $k|n$ . In the following we will prove that the set  $\mathcal{U}_n$  is a dense subset. We need several lemmas before giving the rest of the proof.

Let  $P_n^*(\Gamma)$  be those periodic points in  $P_n(\Gamma)$  with minimal period less than  $n$ , and  $\bar{P}_n(\Gamma)$  be those with period exactly equal  $n$ . We deal with these two parts separately. Although a periodic point in  $P_k(\Gamma)$  for  $\Gamma \in \mathcal{U}_k$  is non-degenerate under  $F^k$ , it may be degenerate under  $F^n$ .

**Lemma 4.4.** *Let  $k < n$  with  $k|n$ . Then there is an open and dense subset  $\mathcal{U}_{k,n} \subset \mathcal{U}_k$ , such that for each  $\Gamma \in \mathcal{U}_{k,n}$ , all periodic points in  $P_k(\Gamma)$  are non-degenerate under  $F^n$ .*

*Proof.* It follows from the definition that  $\mathcal{U}_{k,n}$  is open in  $\mathcal{U}_k$ . So we only need to show the denseness of  $\mathcal{U}_{k,n}$  in  $\mathcal{U}_k$ . Pick  $\Gamma \in \mathcal{U}_k \cap \mathcal{S}_k$ . Then we perturb one reflection point on each periodic orbit  $\mathcal{O}(x)$ , say  $\Gamma_{\epsilon, x}$  such that the rotation number  $\rho_{\epsilon}(x)$  of that orbit changes (see Lemma 3.1). Note that the new rotation number depends continuously on the size of the perturbation. By choosing  $\epsilon(x)$  properly, we can assume the new rotation number is irrational. Note that  $|P_k(\Gamma)|$  is locally constant and  $P_k(\Gamma)$  varies continuously with respect to  $\Gamma \in \mathcal{U}_k$ . After a finite steps of perturbations, the new table is in  $\mathcal{U}_{k,n}$ .  $\square$

Let  $\beta(x) = \max\{|\pi_2(F^i x) - \pi/2| : 0 \leq i < n\}$  for each  $x \in \bar{P}_n(\Gamma)$ , and  $\beta_n(\Gamma) = \inf\{\beta(x) : x \in \bar{P}_n(\Gamma)\}$ .

**Lemma 4.5.** *Let  $\mathcal{U} = \bigcap_{k < n: k|n} \mathcal{U}_{k,n}$ , and  $\Gamma \in \mathcal{U}$ . Then  $\beta_n(\Gamma) > 0$ .*

*Proof.* Note that if  $\beta(x) = 0$ , then  $x$  must be a periodic point of period 2. So  $\beta(x) > 0$  for each  $x \in \bar{P}_n(\Gamma)$ , since  $n \geq 3$ . Suppose on the contrary that there exists  $x_k \in \bar{P}_n(\Gamma)$  with  $\beta(x_k) \rightarrow 0$ . Passing to a subsequence if necessary, we assume  $x_k \rightarrow x$ , which implies  $F^n x = x$  and  $x$  is degenerate under  $F^n$ . Moreover, we have  $s(x) = 0$ , which implies that  $x$  is of period 2. This is impossible if  $n$  is odd. If  $n$  is even, it contradicts our choice of  $\mathcal{U} \subset \mathcal{U}_{k,n}$  for  $k = 2$ .  $\square$

**Remark 4.** One advantage for the proof of the denseness of  $\mathcal{U}_2$  is that all periodic orbits of period 2 move in one direction  $\theta = \pi/2$ . So we can shift  $I_i^r$  in any direction different from  $\pi/2$  (we chose  $\pi/4$  in the proof). For periodic orbits of higher periods, it may not be true that one can find a uniform direction that are transverse to all periodic trajectories. Lemma 4.5 guarantees that, for  $n \geq 3$ , one can always shift along the direction  $\theta \in (\pi/2 - \beta_n(\Gamma), \pi/2)$  at *some iterate* of a periodic orbit on  $\bar{P}_n(\Gamma)$ . This is sufficient for our construction of perturbations.

*Proof of Proposition 4.2: the denseness of  $\mathcal{U}_n$  for  $n \geq 3$ .* Let  $\mathcal{U}_{k,n} \subset \mathcal{U}_k$  be given by Lemma 4.4, and  $\mathcal{U} = \bigcap_{k < n: k|n} \mathcal{U}_{k,n}$ . It suffices to show that  $\mathcal{U}_n$  is dense in  $\mathcal{U}$ . Now let  $\Gamma \in \mathcal{U} \cap \mathcal{S}_n$ , where  $\mathcal{S}_n$  is given by Proposition

2.2. Then every periodic orbit in  $P_n(\Gamma)$  has zero defect.

It is important to notice that, each periodic point  $x \in P_n^*(\Gamma)$  is nondegenerate under  $F^n$  (since we choose  $\Gamma \in \mathcal{U}$ ), and isolated in  $P_n(\Gamma)$ . So we can pick an open neighborhood  $U \supset P_n^*(\Gamma)$ , such that  $P_n(\hat{\Gamma}) \cap \bar{U} = P_n^*(\hat{\Gamma}) \subset U$  for all  $\hat{\Gamma}$  close to  $\Gamma$ . Then the graph of  $F^n$  is transverse to  $\Delta$  along  $\bar{U}$  for all nearby  $\hat{\Gamma}$ , and we only need to consider the part  $M_n \setminus U$ .

Let  $x \in \bar{P}_n(\Gamma) := P_n(\Gamma) \setminus P_n^*(\Gamma)$  be a point of period  $n$ . Then the set  $\pi_1(\mathcal{O}(x))$  consists  $n$  reflection points on  $\Gamma$ . Let  $s(x)$  be the minimal separation of  $\pi_1(\mathcal{O}(x))$  on  $\Gamma$ . Clearly  $s(x) > 0$  for each  $x \in \bar{P}_n(\Gamma)$ .

**Claim.**  $s_n(\Gamma) = \inf\{s(x) : x \in \bar{P}_n(\Gamma)\} > 0$ .

*Proof.* Suppose on the contrary that there exists  $x_k \in \bar{P}_n(\Gamma)$  with  $s(x_k) \rightarrow 0$ . Passing to a subsequence if necessary, we assume  $x_k \rightarrow x$ , which implies  $F^n x = x$  and  $x$  is degenerate under  $F^n$ . Since every periodic point in  $P_n^*(\Gamma)$  is nondegenerate under  $F^n$ , we must have  $x \in \bar{P}_n(\Gamma)$  with  $s(x) = 0$ . So the orbit of  $x$  has positive defect, contradicts the choice of  $\Gamma \in \mathcal{S}_n$ .  $\square$

Let  $s_n(\Gamma)$  be given as above, and  $\epsilon > 0$  be a positive number. Pick a sequence of open intervals  $I_i = (s_i - \epsilon \cdot s_n(\Gamma), s_i + \epsilon \cdot s_n(\Gamma))$  such that the cores  $I_i^r = (s_i - \epsilon^r \cdot s_n(\Gamma), s_i + \epsilon^r \cdot s_n(\Gamma))$  cover  $\Gamma$ . Then we can

cover  $M_n \setminus U$  by much smaller balls  $\{B_j : j = 1, \dots, m_n\}$  such that  $\pi_1(F^k B_j) \subset I_i^r$  (for some  $i = i(k, j)$ ), for each  $k = 0, \dots, n$ .

For each  $x \in \bar{P}_n(\Gamma)$ , we have

- (1)  $|\pi_2(F^k x) - \pi/2| \geq \beta_n(\Gamma)$  for some  $k$ , if the orbit of  $x$  is not symmetric;
- (2)  $\pi_2(F^k x) = \pi/2$  for some  $k$ , if the orbit of  $x$  is symmetric.

Then the perturbation below will be made at the reflection point of  $F^k x$ . Note that if the graph of  $F^n$  is transverse to  $\Delta$  at some  $F^k x$ , then it is transverse to  $\Delta$  along the whole orbit  $\mathcal{O}(x)$ .

Without loss of generality, we assume  $k = 0$ . The perturbations we need here are similar to those we used for proving the denseness of  $\mathcal{U}_2$ , just here we fix a direction  $\theta \in (\pi/2 - \beta_n(\Gamma), \pi/2)$ . Then the perturbation  $\Gamma_i(s, \alpha)$  is supported on  $I_i$  that shifts the core part  $I_i^r$  along the  $\theta$  direction, and then rotates the tangent direction. There is an open neighborhood  $D_i$  of  $(0, 0)$  such that  $\Gamma_i(s, \alpha) \in \Upsilon^r(S^2, g)$ . Note that for each  $x \in \bar{P}_n(\Gamma)$ , there exists an  $i$  such that  $\pi_1(x) \in I_i^r$ , while  $\pi_1(F^k x) \notin I_i$  (by the choice of  $I_i$ ) for all  $k = 1, \dots, n-1$ . Then  $F_{i,s,\alpha} \equiv F$  for all  $(s, \alpha)$  sufficiently small on the set of points not based on  $I_i$ . In particular,  $F_{i,s,\alpha}^k(x) = F^{k-1} \circ F_{i,s,\alpha}(x)$  for all  $k \geq 1$  till the first return of  $x$  to  $I_i$ , and  $\{F_{i,s,\alpha}^k(x) : (s, \alpha) \in D_i\} = F^{k-1}(\{F_{i,s,\alpha}(x) : (s, \alpha) \in D_i\})$  contains an open neighborhood of  $F^k(x)$  for all such  $k$ . Therefore, the evaluation map  $\zeta_i^{\text{ev}} : D_i \times M \rightarrow M \times M, (s, \alpha, x) \mapsto (x, F_{i,s,\alpha}^n(x))$  is transverse to  $\Delta$  along  $(0, 0) \times B_i$ . Putting these together, we obtain a map

$$\zeta^{\text{ev}} : D_1 \times \dots \times D_{m_n} \times M \rightarrow M \times M, ((s_i, \alpha_i)_{i=1}^{m_n}, x) \mapsto (x, F_{(s_i, \alpha_i)}^n(x)),$$

which is transverse to  $\Delta$  along  $0^{2m_n} \times \bigcup_j B_j \supset 0^{2m_n} \times M_n \setminus \bar{U}$ . By the openness properties of transverse intersection, there is an open neighborhood  $D$  of  $0^{2m_n}$ , and a residual subset of parameters  $(s_j, \alpha_j)$  in  $D$  such that the graph of  $F_{(s_j, \alpha_j)}^n$  is transverse to  $M_n \setminus U$ , and hence on  $M$ . In particular,  $\Gamma$  can be approximated by  $\hat{\Gamma}$  in  $\mathcal{U}_n$ . This prove the denseness of  $\mathcal{U}_n$  in  $\mathcal{U}$  and hence in  $\Upsilon^r(S^2, g)$ .  $\square$

In the previous part of this section, we fix the regularity  $r \geq 2$  and use the notation  $\mathcal{U}_n$ . Now we switch to  $\mathcal{U}_n^r$  to indicate the dependence of  $\mathcal{U}_n$  on  $r$ . Let  $\mathcal{R}^r = \bigcap_{n \geq 2} \mathcal{U}_n^r$ . A periodic point is said to be *elementary*, if it is either hyperbolic, or elliptic with irrational rotation number.

**Theorem 4.** *There exists a residual subset  $\mathcal{R}^r$  of  $\Upsilon^r(S^2, g)$ , such that for each  $\Gamma \in \mathcal{R}^r$ , every periodic point of the billiard map induced on  $\Gamma$  is elementary.*

**Remark 5.** The proof of Theorem 4 among the abstract space  $\text{Diff}_\mu^r(M)$  was given in [Rob70]. Robinson's proof is based on some version of transversality theorem. The proof using Parametric Transversality Theorem was given later in his book [Rob95]. Generally speaking, the transversality result applies if the perturbation space is rich enough. This richness is not that obvious in the study of dynamical billiards, since the perturbations of the billiard map  $F$  can only be made via deformations of the billiard table  $Q$ .

Our proof does not apply to the case  $r = \infty$  (at least not directly). The dynamical nature of the property ensures the genericity in  $C^\infty$  category.

**Theorem 5.** *There is a residual subset  $\mathcal{R}^\infty \subset \Upsilon^\infty(S^2, g)$ , such that for each  $\Gamma \in \mathcal{R}^\infty$ , every periodic point of the billiard map induced on  $\Gamma$  is elementary.*

*Proof.* Consider the set  $\mathcal{U}_n^\infty = \left( \bigcup_{r \geq n} \mathcal{U}_n^r \right) \cap \Upsilon^\infty(S^2, g)$ : this set is open in  $\Upsilon^\infty(S^2, g)$  and  $C^r$  dense for each  $r \geq n$ . Therefore  $\mathcal{U}_n^\infty$  is open and dense in  $\Upsilon^\infty(S^2, g)$ . Let  $\mathcal{R}^\infty = \bigcap_{n \geq 2} \mathcal{U}_n^\infty$ .  $\square$

**4.4. Transverse heteroclinic intersections.** Let  $\mathcal{V}_n \subset \Upsilon^r(S^2, g)$  be the set of strictly convex domains  $Q \subset S^2$  such that

- (a). each periodic orbit in  $P_n(\Gamma)$  has zero defect;
- (b). any two periodic orbits in  $P_n(\Gamma)$  has no common reflection points.

Note that the set  $\mathcal{V}_n$  itself may not be open. The following proof is based on our understanding of the properties of the billiard maps in the set  $\mathcal{U}_n$ , which is open and dense.

**Proposition 4.6.** *The set  $\mathcal{V}_n$  contains an open and dense subset of  $\Upsilon^r(S^2, g)$ .*

*Proof.* The denseness follows from Proposition 2.2. It suffices to show the openness of  $\mathcal{V}_n$  in  $\mathcal{U}_n$ . Let  $\pi_1 : M \rightarrow \Gamma$  is the projection to the first coordinate,  $s_n(\Gamma)$  be the minimal separation between the points in  $\pi_1(P_n(\Gamma)) \subset \Gamma$ . Then  $s_n(\Gamma) > 0$  for each  $\Gamma \in \mathcal{U}_n \cap \mathcal{R}_0$ . Pick a small open neighborhood  $\mathcal{U} \subset \mathcal{U}_n$  on which  $|P_n(\cdot)|$  is constant and  $P_n(\cdot)$  varies continuously. Then there exists a smaller neighborhood  $\mathcal{V} \subset \mathcal{U}$  of  $\Gamma$ , such that  $s_n(\hat{\Gamma}) > 0$  for each  $\hat{\Gamma} \in \mathcal{V}_n$ . Therefore,  $\mathcal{V}_n$  is open in  $\mathcal{U}_n$ . This completes the proof.  $\square$

Given a hyperbolic periodic point  $p$ , and its stable and unstable manifolds  $W^{s,u}(p)$ , we let  $W_{\pm}^{s,u}(p)$  be the branches of  $W^{s,u}(p) \setminus \{p\}$ . Let  $\mathcal{W}_n \subset \Upsilon^r(S^2, g)$  be the set of convex domains  $\Gamma \in \Upsilon^r(S^2, g)$ , such that for each pair of hyperbolic periodic points  $p, q \in P_n(\Gamma)$ , either  $W^s(p)_{\pm} \cap W_{\pm}^u(q) = \emptyset$ , or  $W_{\pm}^s(p) \cap_x W_{\pm}^u(q)$  for some  $x \in W_{\pm}^s(p) \cap W_{\pm}^u(q)$ .

**Proposition 4.7.** *The set  $\mathcal{W}_n$  contains an open and dense subset of  $\Upsilon^r(S^2, g)$ .*

To prove this result, we need the following perturbation result of Donnay [Don05].

**Lemma 4.8.** *Let  $\Gamma \in \Upsilon^r(S^2, g)$ . For each  $i = \pm 1$ , let  $x_i = F^i x_0$ ,  $c_i : (-\epsilon, \epsilon) \rightarrow M$  be a smooth curve passing  $c_i(0) = x_i$  such that  $Fc_{-1}$  doesn't focus at  $s_0 = \pi_1(x_0)$ , and is tangent to  $F^{-1}c_1$  at  $x_0$ . Then there is a  $C^r$  small perturbation of  $\Gamma$  at the base point  $s_0$  such that  $\hat{F}c_{-1}$  and  $\hat{F}^{-1}c_1$  are transverse at  $x_0$ .*

*Proof.* We consider the perturbations  $\hat{\Gamma}$  satisfying  $\hat{\Gamma}(s_0) = \Gamma(0)$ ,  $\hat{\Gamma}'(s_0) = \Gamma'(0)$  but  $\hat{\kappa}(s_0) = \kappa(s_0) + \epsilon$ . If the perturbation is localized at  $s_0 = \pi_1(x_0)$ , then one always has  $x_i = \hat{F}^i x_0$ , and hence  $x_0 \in \hat{F}c_{-1} \cap \hat{F}^{-1}c_1$ .

The nonfocusing assumption of  $Fc_{-1}$  means that  $\mathcal{B}^-(DF\dot{c}_{-1}(0)) \neq \infty$ , and tangency assumption means that  $\mathcal{B}^-(DF\dot{c}_{-1}(0)) = \mathcal{B}^-(DF^{-1}\dot{c}_1(0))$ . Suppose  $\hat{\kappa}(s_0) \neq \kappa(s_0)$  after the perturbation. First note that  $\mathcal{B}^-(D\hat{F}\dot{c}_{-1}(0))$  and  $\mathcal{B}^+(D\hat{F}^{-1}\dot{c}_1(0))$  stay unchanged, since these quantities do not depend on the reflection with  $\hat{\Gamma}(s_0)$ . Then according to the Mirror Formula,

$$\mathcal{B}^+(D\hat{F}\dot{c}_{-1}(0)) = \mathcal{B}^-(DF\dot{c}_{-1}(0)) - \frac{2\hat{\kappa}(s_0)}{\sin \theta_0} = \mathcal{B}^+(DF\dot{c}_{-1}(0)) - \frac{2\epsilon}{\sin \theta_0}.$$

Therefore  $m(D\hat{F}\dot{c}_{-1}(0)) = m(DF\dot{c}_{-1}(0)) + \epsilon$ , and the intersection is transverse at  $x_0$ .  $\square$

*Proof of Proposition 4.7.* We will show that  $\mathcal{W}_n$  contains an open and dense subset of  $\mathcal{V}_n$ . Pick a small open set  $\mathcal{V} \subset \mathcal{V}_n$  on which  $|P_n(\cdot)|$  is constant and  $P_n(\cdot)$  is continuous. It suffices to show that  $\mathcal{W}_n$  contains an open and dense subset in every such  $\mathcal{V}$ .

We enumerate  $P_n(\Gamma)$  as  $\{p_i(\Gamma) : 1 \leq i \leq I\}$ . Given  $1 \leq i, j \leq I$ ,  $\alpha, \beta \in \{+, -\}$ , let  $\mathcal{W}_{ij\alpha\beta}$  be those  $\Gamma \in \mathcal{W}$  such that either  $W_{\alpha}^s(p_i) \cap W_{\beta}^u(p_j) = \emptyset$ , or  $W_{\alpha}^s(p_i) \cap_x W_{\beta}^u(p_j)$  for some  $x \in W_{\alpha}^s(p_i) \cap W_{\beta}^u(p_j)$ . It suffices to show each  $\mathcal{W}_{ij\alpha\beta}$  contains an open and dense subset in  $\mathcal{V}$ , since  $\bigcap \{\mathcal{W}_{ij\alpha\beta} : 1 \leq i, j \leq I, \alpha, \beta \in \{+, -\}\}$  is contained in  $\mathcal{W}_n$ . In the following we will fix  $ij$  and  $\alpha\beta$ .

Note that there is a simple dichotomy for  $\Gamma \in \mathcal{V}$ :

- (1) either there exist  $\Gamma_k \rightarrow \Gamma$  such that  $W_{\alpha}^s(p_i(k))$  and  $W_{\beta}^u(p_j(k))$  intersect at some point, say  $x_k$ .
- (2) or there is a smaller neighborhood of  $\Gamma$  among which  $W_{\alpha}^s(p_i)$  and  $W_{\beta}^u(p_j)$  don't intersect.

It suffices to show the ones in the first alternative can be approximated by transverse intersections. From now on we fix  $\Gamma_k$  such that  $W_{\alpha}^s(p_i(k))$  and  $W_{\beta}^u(p_j(k))$  intersect non-transversely at  $x_k$ , and drop the dependence on  $k$  safely.

Note that the minimal separation  $s_n(\Gamma) > 0$ , and the orbit  $F^k x$  approximate  $p_i$  ( $p_j$ ) exponentially fast as  $k \rightarrow +\infty$  (as  $k \rightarrow -\infty$ ). By taking some iterates of  $x$  sufficiently close  $\pi_1(p_i)$  if necessary, we can assume that there exists an open interval  $I \subset \Gamma$  of  $s_0 = \pi_1(x)$  such that all other iterates of  $x$  stay out of  $I$ . Now we consider the wavefront at  $x$  generated by the stable and unstable branches. Note that there is no conjugate point in  $Q$ . So no wavefront can focus at  $x$  and  $fx$  simultaneously. Without loss of generality we assume they don't focus at  $x$ . Then we can make a very small perturbation of  $\Gamma$  supported on  $I$ , such that  $W_{\alpha}^s(p_i)$

and  $W_\beta^u(p_j)$  intersect transversely at  $x$  (see Lemma 4.8). Note that transverse intersection, once created, is robust under perturbations. Therefore  $\mathcal{W}_{i\alpha\beta}$  contains an open and dense subset in  $\mathcal{V}$ . This completes the proof.  $\square$

Let  $\mathcal{R}_{KS}^r = \bigcap_{n \geq 2} \mathcal{W}_n$ , which contains a residual subset of  $\Upsilon^r(S^2, g)$ .

**Theorem 6.** *There is a residual subset  $\mathcal{R}_{KS}^r$  of  $\Upsilon^r(S^2, g)$ , such that for each  $\Gamma \in \mathcal{R}_{KS}^r$ ,*

- (1) *every periodic point of  $F$  is elementary;*
- (2) *for any two hyperbolic branches  $W_\alpha^s(p)$  and  $W_\beta^u(q)$ ,*
  - (2a) *either  $W_\alpha^s(p) \cap W_\beta^u(q) = \emptyset$ ,*
  - (2b) *or  $W_\alpha^s(p) \cap_x W_\beta^u(q)$  for some  $x \in W_\alpha^s(p) \cap W_\beta^u(q)$ .*

The case  $r = \infty$  can be obtained in the same way as we did for Theorem 5.

**Remark 6.** This resembles the Kupka–Smale properties for convex billiards. However, the above theorem does not claim  $W_\alpha^s(p)$  and  $W_\beta^u(q)$  are transverse, neither that  $W_\alpha^s(p)$  and  $W_\beta^u(q)$  have nontrivial intersection. In general,  $W_\alpha^s(p)$  and  $W_\beta^u(q)$  may be separated by some (KAM) invariant curves, and this separation is persistent under perturbations. In next section we will study the case when  $p = q$ .

## 5. HOMOCLINIC INTERSECTIONS FOR HYPERBOLIC PERIODIC POINTS

In this section we study the existence of homoclinic intersections of hyperbolic periodic points of convex billiards on  $(S^2, g)$ . Our main result is the following.

**Proposition 5.1.** *There is an open and dense subset  $\mathcal{X}_n \subset \Upsilon^r(S^2, g)$  such that for each  $\Gamma \in \mathcal{X}_n$ , there exist transverse homoclinic intersections for each hyperbolic periodic point  $p \in P_n(\Gamma)$ .*

It suffices to show such  $\mathcal{X}_n$  is open and dense in  $\mathcal{W}_n$  (see Proposition 4.7 for the set  $\mathcal{W}_n$ ). Note that  $P_n(\Gamma)$  is finite and depends continuously for  $\Gamma \in \mathcal{V}_n$ , and the existence of transverse intersections is an open condition. Then  $\mathcal{X}_n$  is automatically open in  $\mathcal{W}_n$ . So it suffices to show the denseness of  $\mathcal{X}_n$  in  $\mathcal{W}_n$ , and it suffices to show the denseness for  $r = \infty$  (see Theorem 5 and the one-sentence remark right after Theorem 6). So we will assume  $r = \infty$  for during the proof of denseness of  $\mathcal{X}_n$ . We need some preparations before giving the proof. We first cut off the relations between the elliptic periodic points and the hyperbolic periodic points of  $F$ .

**5.1. Stability of elliptic periodic points.** Let  $f \in \text{Diff}_\mu^\infty(M)$  and  $p$  be a fixed point of  $f$ . Then  $p$  is said to be *elliptic* if the eigenvalue of  $D_x f : T_x M \rightarrow T_x M$  satisfies  $\lambda \neq 1$  and  $|\lambda_p| = 1$ . An elliptic fixed point  $p$  is said to be (nonlinearly) *stable*, if there are nesting closed disks  $\{D_n\}$  with  $p \in D_{n+1} \subset D_n^o$  such that  $\bigcap_n D_n = \{p\}$  and  $f|_{\partial D_n}$  is transitive. Note that stable fixed points are isolated from the dynamics, and the invariant rays either coincide with  $\partial D_n$ , or are disjoint with  $\partial D_n$ .

Moser proved in [Mos73] his Twist Map Theorem, which says that an elliptic fixed point  $p$  is stable, if there exists  $n \geq 1$  such that the eigenvalue of  $D_p f$  satisfies  $\lambda_p^i \neq 1$  for each  $1 \leq i \leq q$ , and  $a_j(f^n, p) \neq 0$  for some  $1 \leq j \leq [n/2] - 1$ , where  $a_k, k \geq 1$  are the coefficients of Birkhoff normal form around  $p$ . In this case,  $p$  is also said to be Moser stable. By perturbing the Birkhoff normal form and then applying Moser twist map theorem, Robinson proved in [Rob70] that generically, each elliptic periodic point is Moser stable.

It is expected that a small perturbation of the billiard table will change the coefficients of Birkhoff normal form, and make an elliptic periodic point stable. However, it is quite difficult (if not impossible) to compute the Birkhoff normal form for convex billiard dynamics on a convex sphere with non-constant curvature, since we don't know too much about the explicit form around an elliptic periodic point, and the dependence of  $a_k(f^n, p)$  is quite involved (see [DOP03, BuGr10] for the planar case).

In the following we will take a different (simpler) approach to improve the stability of an elliptic periodic points. For an elliptic periodic point  $p$ , the rotation number  $\rho$  of  $p$  is given by the rotation number of projective action  $[D_p F^n]$  on the projective space  $\mathbb{P}^1$ . Then  $p$  is said to have Diophantine rotation number, if  $\rho$  is Diophantine. That is, there exists positive numbers  $c, \tau$  such that

$$\left| \rho - \frac{p}{q} \right| \geq \frac{c}{|q|^{2+\tau}}, \text{ for all rational numbers } \frac{p}{q}. \quad (5.1)$$

Herman proved that the invariant curves with Diophantine rotation numbers are not isolated. See [Yoc92] for details.

**Proposition 5.2.** *Let  $f \in \text{Diff}_\mu^\infty(M)$  and  $p$  be an elliptic fixed point of  $f$  with rotation number  $\rho$ . If  $\rho$  is Diophantine, then  $p$  is stable.*

The main idea of the proof is, by adding a parameter  $\gamma$  to the original system  $f$ , one can create an *artificial* nondegenerate twisting condition for  $f_\gamma$ . Then KAM theory applies to this family  $f_\gamma$ . Although most of the invariant curves are not related to the initial map  $f$ , there do exist a family of invariant curves accumulating to the fixed point with zero twist. These curves are invariant under  $f_0 = f$ . This technique was developed independently by Xia in [Xia92], where he proved the persistence of  $n$ -tori in  $(n+1)$ -dimensional systems. See also [HX13] for some applications of Proposition 5.2 to Lagrangian equilibrium solutions of circular restricted three body problem.

**Proposition 5.3.** *There is a dense subset  $\mathcal{D}_n \subset \mathcal{W}_n$  such that for each  $\Gamma \in \mathcal{D}_n$ , all elliptic periodic points in  $P_n(\Gamma)$  are stable.*

*Proof.* Given a convex domain  $\Gamma \in \mathcal{W}_n$ , pick a sufficiently small neighborhood  $\mathcal{U} \subset \mathcal{W}_n$  of  $\Gamma$  such that  $P_n : \hat{\Gamma} \in \mathcal{U} \mapsto P_n(\hat{\Gamma})$  has the same (finite) cardinality and varies continuously. Note that each periodic point  $p \in P_n(\Gamma)$  has zero defect. We make a small perturbation of  $\Gamma$  around one point  $p$  from each elliptic periodic orbit  $\mathcal{O}(p)$  in  $P_n(\Gamma)$ , say the resulting domain  $\hat{\Gamma}(\epsilon)$ , such that the rotation number  $\rho_\epsilon$  of  $p$  respecting the billiard map on  $\hat{\Gamma}(\epsilon)$  is different from the initial rotation number, see Proposition 3.1. Note that the set of Diophantine numbers has full measure on the interval  $(\rho, \rho_\epsilon)$ . Picking a smaller size if necessary, we can assume  $\rho_\epsilon$  is already Diophantine.

Any two periodic orbits in  $P_n(\Gamma)$  have no common reflection points. So the above perturbation can be localized at one reflection point and they have disjoint supports on  $\Gamma$ . In particular the Diophantine rotation numbers of the already perturbed ones are preserved by the subsequent perturbations.

After a finite steps (at most  $|P_n(\Gamma)|$ ) of perturbations, we arrive at some  $\hat{\Gamma} \in \mathcal{U}$  such that  $P_n(\hat{\Gamma}) = P_n(\Gamma)$ ,  $\hat{F} = F$  on  $P_n(\hat{\Gamma})$  and  $\rho(p, \hat{F})$  is Diophantine for each  $p \in P_n(\hat{\Gamma})$ . Then Proposition 5.2 guarantees that each elliptic periodic point in  $P_n(\hat{\Gamma})$  is stable. Such a perturbation  $\hat{\Gamma}$  can be made arbitrarily close to  $\Gamma$ . Therefore,  $\mathcal{D}_n$  is dense in  $\mathcal{W}_n$ .  $\square$

**5.2. Homoclinic intersections.** Now we study the hyperbolic periodic points in  $P_n(\Gamma)$ . Although each point  $x \in P_n(\Gamma)$  is fixed by  $F^n$ , the two branches of the stable (and unstable) manifolds  $x$  may be switched by  $F^n$ . However,  $F^{2n}$  does fix each branch of the invariant manifolds of hyperbolic periodic points in  $P_n(\Gamma)$ . When studying  $P_n(\Gamma)$ , we actually consider the  $2n$ -th iteration  $F^{2n}$  of those  $\Gamma \in \mathcal{D}_{2n}$ . For simplicity we denote  $f = F^{2n}$ .

Let  $L$  is a branch of the unstable manifold  $W^u(p) \setminus \{p\}$ . Then for any  $x \in L$ , the segment  $L[x, fx]$  can be viewed as a fundamental domain of  $L$  with respect to  $f = F^{2n}$ . As  $k \rightarrow +\infty$ ,  $f^{-k}L[x, fx]$  converges to  $p$ , while  $f^kL[x, fx]$  may have various limiting behaviors. Denote by  $\omega(L)$  the limit set of  $f^kL[x, fx]$  as  $k \rightarrow +\infty$ . Similarly we define the  $\omega$ -set<sup>2</sup> of stable branches (with respect to  $f^{-k}$ ). There is a dichotomy for the branches of invariant manifolds (see [Oli87]):

- either  $\omega(L) \supset L$ , or  $\omega(L) \cap L = \emptyset$ .

A stronger dichotomy was obtained in [XZ14].

**Proposition 5.4.** *Let  $f \in \text{Diff}_\mu(M)$  such that each fixed point is nondegenerate, and each elliptic fixed point is stable. Let  $L$  be a branch of invariant manifolds of a hyperbolic fixed point  $p$ . Assume  $fL = L$ . Then*

- either  $\omega(L) \supset L$ ,
- or  $\omega(L) = \{q\}$  is a singleton, where  $q$  is a hyperbolic fixed point.

The branch  $L$  with  $\omega(L) = \{q\}$  is called a saddle connection. A saddle connection is said to be a homoclinic (heteroclinic, respectively) connection if  $q = p$  ( $q \neq p$ , respectively).

<sup>2</sup>Technically, one should say the  $\alpha$ -set of a stable branch. We use the same notation for stable and unstable branches just to unify the presentation of this paper.

*Proof.* We sketch the main idea of the proof. See [XZ14] for details. Let  $L$  be a branch of the unstable manifold of  $p$ . Suppose  $\omega(L) \not\supset L$ . Then  $\omega(L) \cap L = \emptyset$ . Let  $K$  be the closure of  $L$ , and  $U$  be a connected component of  $M \setminus K$  attached to  $L$ . Let  $\hat{U}$  be the prime-end compactification of  $U$ , whose boundary consists of finitely many circles. One of the circles, say  $C_p$ , contains the prime point  $\hat{p}$  of  $p$ . The restriction of  $\hat{f}$  on  $C_p$  is a circle diffeomorphism, and admits  $\hat{p}$  as an expanding fixed point. So there is at least one more point on  $C_p$  fixed by  $\hat{f}$ , say  $\hat{q}$ . Let  $q$  be the underlining point of  $\hat{q}$  on the closure  $\overline{U}$ , which must be fixed by  $f$ . Such a point can't be elliptic, since elliptic ones are stable and can't be approached by invariant curves outside  $D_n$ . Then  $q$  must be a hyperbolic fixed point, and  $L$  forms a branch of the stable manifold of  $q$ . Therefore,  $\omega(L) = \{q\}$ .  $\square$

As a corollary, we obtain the following result due to Mather [Mat81]. Our formulation is slightly stronger. See also [XZ14, Corollary 3.4].

**Corollary 5.** *Let  $f \in \text{Diff}_\mu(M)$  such that each fixed point of  $f$  is either hyperbolic, or elliptic with Diophantine rotation number. Let  $p$  be hyperbolic fixed point such that all four branches of  $W_\pm^{s,u}(p)$  are fixed by  $f$ . Then either one of the branch forms a saddle connection, or all four branches have the same closure.*

*Proof.* Pick a local coordinate system  $(U, (x, y))$  around  $p$  such that the branches leave  $p$  along the two axes. Suppose none of the four branches is a saddle connection. Then each branch is recurrent, and its  $\omega$ -set contains the branch itself and least one of the branches adjacent to it. If the  $\omega$ -set of a branch  $L$  does not contain the other adjacent branch, say  $K$ , then consider the component  $C$  of  $M \setminus \overline{L}$  containing the quadrant between  $L$  and  $K$ . The boundary of  $C \cap U$  consists only of pieces of the branches of the invariant manifolds. This forces  $L$  to be a homoclinic loop, contradicts the hypothesis we started with.  $\square$

**Proposition 5.5.** *Let  $\Gamma \in \mathcal{D}_{2n}$ . Then for each hyperbolic periodic point  $x \in P_n(\Gamma)$ , there exist transverse homoclinic intersections between each branch of the stable manifold and each branch of the unstable manifold of  $x$ .*

The proof mainly use the fact that the (algebraic) intersection number between two simple closed curves on  $M$  must be 0. This kind of arguments also appeared in [Rob73, Pix82, XZ14].

*Proof.* Let  $\Gamma \in \mathcal{D}_{2n}$ ,  $F$  be the induced billiard map on  $M = \Gamma \times (0, \pi)$ . Note that there is no saddle connection between any hyperbolic periodic points in  $P_{2n}(\Gamma)$  (by the definition of  $\mathcal{W}_{2n}$  and the fact that  $\mathcal{D}_{2n} \subset \mathcal{W}_{2n}$ ), and each elliptic periodic point in  $P_{2n}(\Gamma)$  is stable (by Proposition 5.3).

Let  $p$  be a hyperbolic periodic point in  $P_n(\Gamma)$ ,  $L$  be a branch of the unstable manifold of  $p$ , and  $K$  be a branch of the stable manifold of  $p$ . Then both  $L, K$  are fixed by  $F^{2n}$ , are recurrent, and they have the same closure (by Corollary 5). Pick a local coordinate system  $(U, (x, y))$  around  $p$  such that  $L$  leaves  $p$  along the positive  $x$ -axis, and  $L$  approximates  $p$  through the first quadrant. Let  $S_\epsilon = \{(x, y) \in U : 0 < x, y \leq 1, xy \leq \epsilon\}$ , and  $q$  be the first moment on  $L$  that hits  $S_\epsilon$ . Let  $C$  be the closed curve that starts from  $p$ , first travels along  $L$  to the point  $q$ , and then the segment  $\overline{qp}$  from  $q$  to  $p$ . Then  $C$  is a simple closed curve.

Since the closure of  $K$  contains  $L$ ,  $K$  also intersects  $S_\epsilon$ . Let  $\hat{C}$  be the corresponding simple closed curve by closing the first intersection  $\hat{q}$  of  $K$  with  $S_\epsilon$ . Then we see that  $C$  and  $\hat{C}$  cross each other at  $p$ , and the two open segments  $(p, q)$  and  $(p, \hat{q})$  do not intersect (by the entrance–exit analysis, see [Oli87, XZ14]). Clearly  $L(p, q) \cap (p, \hat{q}) = \emptyset$  and  $K(p, \hat{q}) \cap (p, q) = \emptyset$ .

However, the algebraic intersection number between any two closed curves on  $M$  must be 0. So  $C$  and  $\hat{C}$  have to cross each other at some point beside  $p$ , say  $y$ , and that intersection must happen between  $L(p, q)$  and  $K(p, \hat{q})$ . Therefore, there is a homoclinic intersection between  $K$  and  $L$ . The intersection at  $y$  is a topological crossing, but may not be transverse. However, transverse homoclinic intersections do exist, since  $\mathcal{D}_{2n} \subset \mathcal{W}_{2n}$ .  $\square$

Note that no perturbation is needed for the proof of the above proposition.

*Proof of Proposition 5.1.* As we discussed right after stating Proposition 5.1,  $\mathcal{X}_n$  is open in  $\mathcal{W}_n$ . Let  $\mathcal{D}_{2n}$  be the dense subset of  $\mathcal{W}_{2n}$  given by Lemma 5.3. Then Proposition 5.5 shows that  $\mathcal{D}_{2n} \subset \mathcal{X}_n$ . Therefore,  $\mathcal{X}_n$  is open and dense in  $\Upsilon^r(S^2, g)$ . This completes the proof.  $\square$



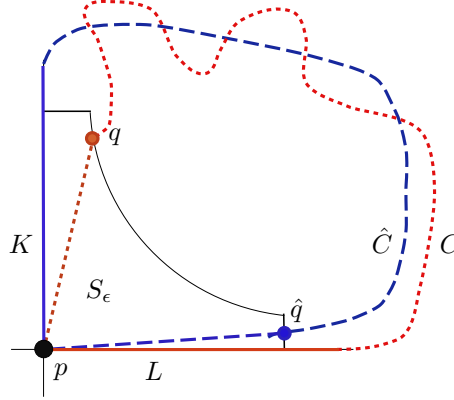


FIGURE 2. The closing curves  $C$  (red) and  $\hat{C}$  (blue) when  $K$  leaves along the positive  $y$ -axis. The case that  $K$  leaves along the negative  $y$ -axis is similar.

*Proof of Theorem 3.* Let  $\mathcal{R}^r = \bigcap_{n \geq 1} \mathcal{X}_n$ . Then  $\mathcal{R}^r$  is a residual subset of  $\Upsilon^r(S^2, g)$ . For each  $f \in \mathcal{R}$ , and each hyperbolic periodic point  $p$ , its stable and unstable manifolds admit some transverse intersections. This completes the proof.  $\square$

The case  $r = \infty$  can be proved in the same way as we did for Theorem 5.

### 5.3. Positive topological entropy.

**Corollary 6.** *There is an open and dense subset  $\mathcal{U} \subset \Upsilon^r(S^2, g)$  such that for each  $\Gamma \in \mathcal{U}$ , the billiard map has transverse homoclinic intersections and positive topological entropy.*

*Proof.* Let  $\mathcal{D}_2$  be the dense subset given in Lemma 5.3. Let  $\Gamma \in \mathcal{D}_2$ . Then each point  $x \in P_2(\Gamma)$  is non-degenerate. Let  $W(s_1, s_2) = S(s_0, s_1) + S(s_1, s_2)$  be the action along the 2-periodic configuration  $(s_k)$  on  $\Gamma$ . Let  $(s_k)$  be a 2-periodic configuration at where  $W$  attains its minimum, and  $x$  the corresponding periodic point of period 2. Then  $D^2W(s_1, s_2)$  is positive definite, and  $\text{Tr}(D_x F^2) > 2$  (see Proposition 2.1). So  $x$  is hyperbolic. Moreover, each branch of the invariant manifolds of  $x$  is fixed by  $F^2$ , since both eigenvalues are positive (the double period iterate  $F^{2n}$  is not needed for minimizers). Then the proof of Proposition 5.5 shows that there exist transverse homoclinic intersections of the stable and unstable branches of  $x$ . Transverse intersections are robust. So there exists an open set  $\mathcal{U} \supset \mathcal{D}_2$  such that each  $\Gamma \in \mathcal{U}$  has transverse homoclinic intersections and positive topological entropy.  $\square$

## APPENDIX A. ZERO DEFECT FOR GENERIC CONVEX BILLIARDS

In this section, we give a proof of Proposition 2.2. Let  $\Upsilon^r(S^2, g) \subset C^r(\mathbb{T}, S^2)$  be the set of convex curves. We will use  $f : \mathbb{T} \rightarrow S^2$  to emphasize the role of  $f$  as an embedding function, and use  $\Gamma = f(\mathbb{T})$  only for its image. Let  $f : \mathbb{T} \rightarrow S^2$  be a simple closed curve enclosing a strictly convex domain  $Q$ ,  $p$  be a nonsymmetric periodic point of the billiard map  $F$  with period  $n = |\mathcal{O}(p)| \geq 3$ . Let  $F^k p = (s_k, \theta_k)$ , and  $\pi_\Gamma(\mathcal{O}(p)) = \{y_1, \dots, y_t\} \subset \Gamma$ . Suppose  $p$  has positive defect:  $d(p) = n - t > 0$ , and  $(y_{w(1)}, \dots, y_{w(n)})$  be the ordered reflection sequence of  $\mathcal{O}(p)$ . This gives rise to an onto map  $w : \{1, \dots, n\} \mapsto \{1, \dots, t\}$ . Such a map  $w$  is said to be the pattern of the orbit  $\mathcal{O}(p)$ . Without loss of generality we assume  $w(1) = 1$  and  $\{1 \leq j \leq n : w(j) = 1\} = \{j_1, \dots, j_r\}$  (with  $j_1 = 1$ ) for some  $r \geq 2$ .

Now let  $\mathcal{O}(p)$  be a symmetric periodic orbit of period  $n$ . Then  $n = 2m$  is an even number, and there are exactly two reflections of right angle with  $\Gamma$ . Suppose  $p$  has positive defect, and  $t$  be the number of distinct reflection points on  $\Gamma$ . Let  $w : \{1, \dots, m, m+1\} \rightarrow \{1, \dots, t\}$  be the pattern of  $\mathcal{O}(p)$  such that  $y_{w(1)}$  and  $y_{w(m+1)}$  are the two reflection points on  $\Gamma$  with right angle. Note that  $w(1) \neq w(m+1)$ . We first study the nonsymmetric case in details. The symmetric case need some minor modifications, and will be given at the end of the proof.

Now we generalize above notations to any closed path of type  $w$  on  $S^2$ . Let  $t \geq 3$  be given. Then a map  $w : \mathbb{Z} \rightarrow \{1, \dots, t\}$  is said to be of period  $n$  if  $w(k+n) = w(k)$  for all  $k$ ; is said to be admissible if  $w(k) \neq w(k+1)$  for all  $k$ . There are only finitely many admissible patterns of period  $n$ , and we will fix such a pattern from now. Let  $\mathbb{T}^{(t)} \subset \mathbb{T}^t$  be the set of points  $(s_1, \dots, s_t)$  with  $s_i \neq s_j$  for all  $i \neq j$ . Then for each  $\mathbf{x} \in \mathbb{T}^{(t)}$  and  $\mathbf{y} = f^{(t)}(\mathbf{x})$ , we have that  $\{y_{w(k)}\}$  is a closed path of type  $w$ .

Let  $\mathbf{y} = (y_1, \dots, y_t)$  be a collection of  $t$  distinct points on  $S^2$ . Define the perimeter of the geodesic polygon with the ordered corners at  $\{y_{w(k)}\}$  as

$$H_w(\mathbf{y}) = \sum_{k=1}^n d(y_{w(k)}, y_{w(k+1)}).$$

Similarly, given  $f : \mathbb{T} \rightarrow S^2$  and  $\mathbf{x} \in \mathbb{T}^{(t)}$ , let  $H_w(f^{(t)}\mathbf{x})$  be the perimeter of the corresponding geodesic polygon with corners  $(f(s_i))$  and pattern  $w$ .

Let  $J^1(\mathbb{T}, S^2)$  be the 1-jet bundle, and  $J_t^1(\mathbb{T}, S^2)$  be the  $t$ -fold jet bundle. For each  $f \in C^r(\mathbb{T}, S^2)$ , we have a section map  $j_t f : \mathbf{x} \in \mathbb{T}^{(t)} \mapsto (jf(s_1), \dots, jf(s_t))$ . Let  $V_w$  be the set of those  $\tau = (jf_1(s_1), \dots, jf_t(s_t))$  such that

- (1)  $f_j(s_j) \neq f_i(s_i)$  for each  $j \neq i$ ,
- (2)  $f'_i(s_i) \neq 0$  for every  $i = 1, \dots, t$ , and
- (3) the polygon generated by  $(f_1(s_1), \dots, f_t(s_t))$  is convex with  $t$  vertices.

Let  $\alpha : J^1(\mathbb{T}, S^2) \rightarrow \mathbb{T}$  be the source map, and  $\beta : J^1(\mathbb{T}, S^2) \rightarrow S^2$  be the target map,  $W := (\alpha^t)^{-1}(\mathbb{T}^{(t)}) \cap V_w$ . Clearly  $W$  is an open submanifold of  $J_t^1(\mathbb{T}, S^2)$ . Given  $\tau \in W$ , there are neighborhoods  $U_i \subset \mathbb{T}$  of  $s_i$  and  $V_i \subset S^2$  of  $f_i(U_i)$  with  $U_i \cap U_j = \emptyset$  and  $V_i \cap V_j = \emptyset$  whenever  $1 \leq i < j \leq t$ , such that

$$\Omega := W \cap \left( \prod_{i=1}^t J^1(U_i, V_i) \right)$$

is an open neighborhood of  $\tau$ . Consider the coordinate map

$$\theta : \Omega \mapsto \prod_{i=1}^t U_i \times T_{V_i} S^2 \simeq \prod_{i=1}^s U_i \times V_i \times \mathbb{R}^2,$$

with  $\theta(\tau) = (\mathbf{u}, \mathbf{v}, A)$ , where  $\mathbf{u} = (u_1, \dots, u_t)$  is the source of  $\tau$ ,  $\mathbf{v} = (v_1, \dots, v_t)$  is the target of  $\tau$ , and  $A = (f'_1(u_1), \dots, f'_i(u_i), \dots, f'_t(u_t)) = \begin{pmatrix} f'_{1,1}(u_1) & \cdots & f'_{i,1}(u_i) & \cdots & f'_{t,1}(u_t) \\ f'_{1,2}(u_1) & \cdots & f'_{i,2}(u_i) & \cdots & f'_{t,2}(u_t) \end{pmatrix}$ .

In the following we separate the role of  $s_1$  from  $s_k$ ,  $2 \leq k \leq t$ . For each  $l = 1, \dots, r$ , let  $a = w(j_l - 1)$  and  $b = w(j_l + 1)$ , and  $\eta_a$  be the tangent direction of the shortest geodesic from  $y_1$  to  $y_a$ , and similarly define  $\eta_b$ . Let  $\mathbf{t}_{y_1} = f'_1(u_1)$  and  $\mathbf{n}_{y_1}$  be the unit tangent and normal directions at  $y_1$ . Then we decompose  $\eta_a + \eta_b$  as

$$\eta_a + \eta_b = \xi_l(\mathbf{y})\mathbf{t}_{y_1} + \zeta_l(\mathbf{y})\mathbf{n}_{y_1},$$

where  $\xi_l(\mathbf{y}) = \langle \eta_a + \eta_b, \mathbf{n}_{y_1} \rangle$  and  $\zeta_l(\mathbf{y}) = \langle \eta_a + \eta_b, \mathbf{t}_{y_1} \rangle$ . Then it follows from the basic properties of billiard maps that

- (1a).  $\zeta_l(\mathbf{y}) = 0$  if  $(y_{w(t)})_{t=1}^n$  is a periodic orbit;
- (1b).  $\zeta_l(\mathbf{y}) \neq 0$  if  $(y_a, y_1, y_b)$  does not describe a reflection.
- (2a).  $\partial_{s_k} H \circ f^{(t)}(\mathbf{x}) = 0$  for orbit paths;
- (2b).  $\partial_{s_k} H \circ f^{(t)}(\mathbf{x})$  may not be zero for non-orbit paths.

Let  $\Sigma_w \subset M$  be those  $\tau = (jf_1(s_1), \dots, jf_t(s_t)) \in M$  so that  $\zeta_l(\mathbf{y}) = 0$  for each  $1 \leq l \leq r$ , and  $\partial_{s_k} F_w \circ (f_1, \dots, f_t)(\mathbf{x}) = 0$  for each  $2 \leq k \leq t$ . We first estimate the codimension of  $\Sigma_w$ . Let  $\tau \in \Sigma_w \subset M$  be given, and  $\theta : \tau \mapsto (\mathbf{u}, \mathbf{v}, A)$  be the coordinate system around  $\tau$  given as above. Define a function

$$\mathcal{K} : \theta(\Omega) \rightarrow \mathbb{R}^{r+t-1}, \quad \chi \mapsto (\phi_1(\chi), \dots, \phi_r(\chi); \psi_2(\chi), \dots, \psi_t(\chi)),$$

where

- (1)  $\phi_l : \theta(\Omega) \rightarrow \mathbb{R}$ ,  $l = 1, \dots, r$ , is defined by

$$\chi = (\mathbf{u}, \mathbf{v}, A) \mapsto \zeta_l(\mathbf{y}) = \langle \eta_a + \eta_b, \mathbf{t}_{y_1} \rangle,$$

- where  $a = w(j_l - 1)$ ,  $b = w(j_l + 1)$ , and  $\mathbf{t}_{y_1}$  is the unit tangent direction along  $f_1(u_1)$ .  
 (2)  $\psi_k : \theta(\Omega) \rightarrow \mathbb{R}$ ,  $\chi \mapsto \langle \nabla_{y_k} H, \mathbf{t}_{y_k} \rangle$ , for each  $k = 2, \dots, t$ .

Note that  $\mathcal{K}(\tau) = 0$  for each  $\tau \in \Sigma_w \cap \Omega$ . We claim that  $\mathcal{K}$  is a submersion at each point in  $\Omega$ . The verification of the submersion is pretty simple for convex billiards: by pushing the point  $y_a$  along the normal direction of  $f_a(s_a)$  (for  $a = w(j_l - 1)$ ), while fixing all other  $y_k$ ,  $k \neq a$ ), we see that  $\phi_l$  changes linearly (since  $\mathbf{t}_{y_1}$  is fixed); by rotating the tangent direction  $\mathbf{t}_{y_k}$  of  $y_k$  along  $f_k(s_k)$ , (while fixing all  $y_k$ ), we see that  $\psi_k$  changes linearly (since  $\nabla_{y_k} F$  is a fixed nonzero vector); and all these variations are independent.

Therefore, the map  $\mathcal{K}$  is a submersion at each point in  $\Omega$ . So the codimension of  $\Sigma_w$  in  $\Omega \subset W$  is at least  $\dim(\text{Im}(\mathcal{K})) = r + t - 1 \geq t + 1$ , which is larger than  $\dim \mathbb{T}^{(t)} = t$ . Then by Multi-jet Transversality Theorem, we have that  $j_t f \cap \Sigma_w = \emptyset$  for a residual subset of convex tables. Similarly, we define  $\Sigma_{w'}$  for any  $n$ -periodic admissible pattern  $w' : \mathbb{Z} \rightarrow \{1, \dots, t\}$ , and then for any  $t = 2, \dots, n - 1$ . This completes the proof for nonsymmetric periodic orbits.

For symmetric periodic orbits, the proof is almost the same. The only difference is that when  $a := w(j_l - 1) = w(j_l + 1)$ , and the collision from  $y_a$  to  $y_1$  is at the right angle. In this case, we still have that  $\phi_l$  changes linearly by pushing  $y_a$  along the normal direction of  $f_a(s_a)$  (since  $\mathbf{t}_{y_1}$  is fixed). Then the rest of the proof is the same. Putting together these results, we get that, for a residual subset of convex tables, each periodic orbits with period  $n$  has zero defect. This completes the proof that genericity of zero defect.

For the second part of Proposition 2.2, we note that in the proof given above, we used the property that each folding of the path at  $y_{w(k)}$  is a reflection; but we didn't use any property that  $\{y_{w(k)}\}$  is on a single orbit. In particular, one can take the union of the two periodic orbits and then study the paths with that joint pattern. Therefore the same analysis applies to the case that two orbits have some common reflection point. Then we conclude that, there is a residual subset of convex tables, for which any two periodic orbits with no common geodesic segment has no common reflection point. However, note that the orbit obtained by the time-reversal of one orbit has exact the same geodesic segments, and this does not count as positive defects.

#### ACKNOWLEDGMENTS

The author thanks Saša Kocić, Jeff Xia and Hong-Kun Zhang for many suggestions and useful discussions.

#### REFERENCES

- [Ban88] V. Bangert. *Mather sets for twist maps and geodesics on tori*. Dynamics Reported **1**, 1–56, Wiley, Chichester, 1988.
- [Bia93] M. Bialy. *Convex billiards and a theorem by E. Hopf*. Math. Z. **214** (1993), 147–154.
- [Bia13] M. Bialy. *Hopf Rigidity for convex billiards on the hemisphere and hyperbolic plane*. Discrete Conti. Dynam. Syst. **33** (2013), 3903–3913.
- [Bir17] G. Birkhoff. *Dynamical systems with two degrees of freedom*. Trans. Amer. Math. Soc. **18** (1917), 199–300.
- [Bir27] G. Birkhoff. *Dynamical systems*. Amer. Math. Soc. Colloq. Pub. **IX**, Providence, R.I. 1966 (original 1927).
- [Bol92] S. V. Bolotin. *Integrable billiards on surfaces of constant curvature*. (Russian) Mat. Zametki **51** (1992), 20–28; (translation) Math. Notes **51** (1992), 117–123.
- [Bun74] L. A. Bunimovich. *On ergodic properties of certain billiards*. Funct. Anal. Appl. **8** (1974), 254–255.
- [Bun92] L. A. Bunimovich. *On absolutely focusing mirrors*. In “Ergodic theory and related topics”, 62–82, Lecture Notes in Math. **1514**, Springer, Berlin, 1992.
- [BuGr10] L. Bunimovich, A. Grigo. *Focusing components in typical chaotic billiards should be absolutely focusing*. Comm. Math. Phys. **293** (2010), 127–143.
- [BZZ14] L. Bunimovich, H-K. Zhang, P. Zhang. *On another edge of defocusing: hyperbolicity of asymmetric lemon billiards*. arXiv:1412.0173 [math.DS].
- [BuGe89] K. Burns, M. Gerber. *Real analytic Bernoulli geodesic flows on  $S^2$* . Ergod. Th. Dynam. Sys. **9** (1989), 27–45.
- [ChMa06] N. Chernov, R. Markarian. *Chaotic billiards*. Math. Surv. Monogr. **127**, AMS, Providence, RI, 2006.
- [CP14] L. Coutinho dos Santos, S. Pinto-de-Carvalho. *Oval Billiards on Surfaces of Constant Curvature*. preprint, 2014.
- [DOP03] M. J. Dias Carneiro, S. Oliffson Kamphorst, S. Pinto-de-Carvalho. *Elliptic islands in strictly convex billiards*. Ergod. Th. Dynam. Syst. **23** (2003), 799–812.
- [DOP07] M. J. Dias Carneiro, S. Oliffson Kamphorst, S. Pinto-de-Carvalho. *Periodic orbits of generic oval billiards*. Nonlinearity **20** (2007), 2453–2462.
- [Don89] V. Donnay. *Geodesic flow on the two-sphere, Part I: positive measure entropy*. Ergod. Th. Dynam. Sys. **8** (1989), 531–553.
- [Don91] V. Donnay. *Using integrability to produce chaos: billiards with positive entropy*. Commun. Math. Phys. **141** (1991), 225–257.

- [Don05] V. Donnay. *Creating transverse homoclinic connections in planar billiards*. J. Math. Sci. **128** (2005), 2747–2753.
- [FLBP] C. L. Foden, M. L. Leadbeater, J. H. Burroughes, M. Peper. *Quantum magnetic confinement in a curved two-dimensional electron gas*. J. Phys. Condens. Matter **6** (1994), L127.
- [FL03] J. Franks, P. Le Calvez. *Regions of instability for non-twist maps*. Ergod. Th. Dynam. Syst. **23** (2003), 111–141.
- [HX13] Y. Hua, Z. Xia. *Stability of elliptic periodic points with an application to Lagrangian equilibrium solutions*. Qual. Theory Dyn. Syst. **12** (2013), 243–253.
- [GSG99] B. Gutkin, U. Smilansky, E. Gutkin. *Hyperbolic Billiards on surfaces of constant curvature*. Comm. Math. Phys. **208** (1999), 65–90.
- [Ivr80] V. Ivrii. *Second term of the spectral asymptotic expansion of the Laplace-Beltrami operator on manifolds with boundary*. Func. Anal. Appl. **14** (1980), 98–106.
- [KSS89] A. Kramli, N. Simanyi, D. Szasz. *Dispersing billiards without focal points on surfaces are ergodic*. Comm. Math. Phys. **125** (1989), 439–457.
- [Laz81] V. F. Lazutkin. *Convex billiards and eigenfunctions of the Laplace operator* (Russian). Leningrad. Univ., Leningrad, 1981.
- [MM83] R. Mackay, J. Meiss. *Linear stability of periodic orbits in Lagrangian systems*. Phys. Lett. A **98** (1983), 92–94.
- [Mar88] R. Markarian. *Billiards with Pesin region of measure one*. Comm. Math. Phys. **118** (1988), 87–97.
- [Mat81] J. Mather. *Invariant subsets of area-preserving homeomorphisms of surfaces*. Adv. Math. Suppl. Stud. **7B** (1981) 531–62.
- [Mat82] J. Mather. *Topological proofs of some purely topological consequences of Caratheodory’s theory of prime ends*. In: Selected Studies, Eds. Th. M. Rassias, G. M. Rassias, 1982, 225–255.
- [Mos73] J. Moser. *Stable and Random Motions in Dynamical Systems*. Annals Math. Studies **77**, Princeton, NJ, 1973.
- [Oli87] F. Oliveira. *On the generic existence of homoclinic points*. Ergod. Th. Dynam. Syst. **7** (1987), 567–595.
- [Oli00] F. Oliveira. *On the  $C^\infty$  genericity of homoclinic orbits*. Nonlinearity **13** (2000), 653–662.
- [PeSt87a] V. Petkov, L. Stojanov. *Spectrum of the Poincar map for periodic reflecting rays in generic domains*. Math. Z. **194** (1987), 505–518.
- [PeSt87b] V. Petkov and L. Stojanov, *Periods of multiple reflecting geodesics and inverse spectral results*. Amer. J. Math. **109** (1987), 619–668.
- [PeSt88] V. Petkov and L. Stojanov, *On the number of periodic reflecting rays in generic domains*. Ergod. Th. Dynam. Syst. **8** (1988), 81–91.
- [Pix82] D. Pixton. *Planar homoclinic points*. J. Differential Equations **44** (1982), 365–382.
- [Poin] H. Poincare. *Les methodes nouvelles de la mecanique celeste*. (French) [New methods of celestial mechanics.] Gauthier-Villars, Paris, vol. 1 in 1892; vol 2 in 1893; vol. 3 in 1899.
- [Pug67] C. Pugh. *The closing lemma*. Amer. J. Math. **89** (1967), 956–1021.
- [PuRo83] C. Pugh, C. Robinson. *The  $C^1$  closing lemma, including Hamiltonians*. Ergod. Theor. Dyn. Sys. **3** (1983), 261–313.
- [Rob70] C. Robinson. *Generic properties of conservative systems*. Amer. J. Math. **92** (1970) 562–603.
- [Rob73] C. Robinson. *Closing stable and unstable manifolds in the two-sphere*. Proc. Am. Math. Soc. **41** (1973), 299–303.
- [Rob95] C. Robinson. *Dynamical systems. Stability, symbolic dynamics, and chaos*. Stud. Adv. Math., CRC Press, 1995.
- [Sar11] P. Sarnak. *Recent progress on the quantum unique ergodicity conjecture*. Bull. Amer. Math. Soc. **48** (2011), 211–228.
- [Sin70] Ya. G. Sinai. *Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards*. Russian Math. Surveys **25** (1970), 137–189.
- [Sto87] L. Stojanov. *Generic properties of periodic reflecting rays*. Ergod. Th. Dynam. Syst. **7** (1987), 597–609.
- [Tak72] F. Takens. *Homoclinic points in conservative systems*. Invent. Math. **18** (1972), 267–292.
- [Vet84] A. Vetier. *Sinai billiard in potential field (construction of stable and unstable fibers)*. in “Limit theorems in probability and statistics”, 1079–1146, Colloq. Math. Soc. Janos Bolyai, 36, North-Holland, Amsterdam, 1984.
- [Vis12] D. Visscher. *Franks’ Lemma in Geometric Contexts*. Thesis, Northwestern University. 2012.
- [Woj86] M. Wojtkowski. *Principles for the design of billiards with nonvanishing Lyapunov exponents*. Comm. Math. Phys. **105** (1986), 391–414.
- [Xia92] Z. Xia. *The existence of the invariant tori in the volume preserving diffeomorphisms*. Ergod. Th. Dynam. Syst. **12** (1992), 621–631.
- [Xia96] Z. Xia. *Homoclinic points in symplectic and volume-preserving diffeomorphisms*. Comm. Math. Phys. **177** (1996), 435–449.
- [Xia06] Z. Xia, *Homoclinic points for area-preserving surface diffeomorphisms*. arXiv:math/0606291.
- [XZ14] Z. Xia, P. Zhang. *Homoclinic points for convex billiards*. Nonlinearity **27** (2014), 1181–1192.
- [Yoc92] Y. Yoccoz. *Travaux de Herman sur les tores invariants*. [Works of Herman on invariant tori] Seminaire Bourbaki, Asterisque **206** (1992), 311–344.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF HOUSTON, HOUSTON, TX 77004.

E-mail address: pzhang@math.uh.edu